THÈSES

PRÉSENTÉES

A LA FACULTÉ DES SCIENCES
DE L'UNIVERSITÉ DE PARIS

POUR OBTENIR

LE GRADE DE DOCTEUR ÈS SCIENCES MATHÉMATIQUES

PAR

M. RAMAKRISHNA SARMA - Yadavalli

PREMIÈRE THÈSE

Sur les tests et sur l'estimation
de paramètres pour certains processus stochastiques
stationnaires

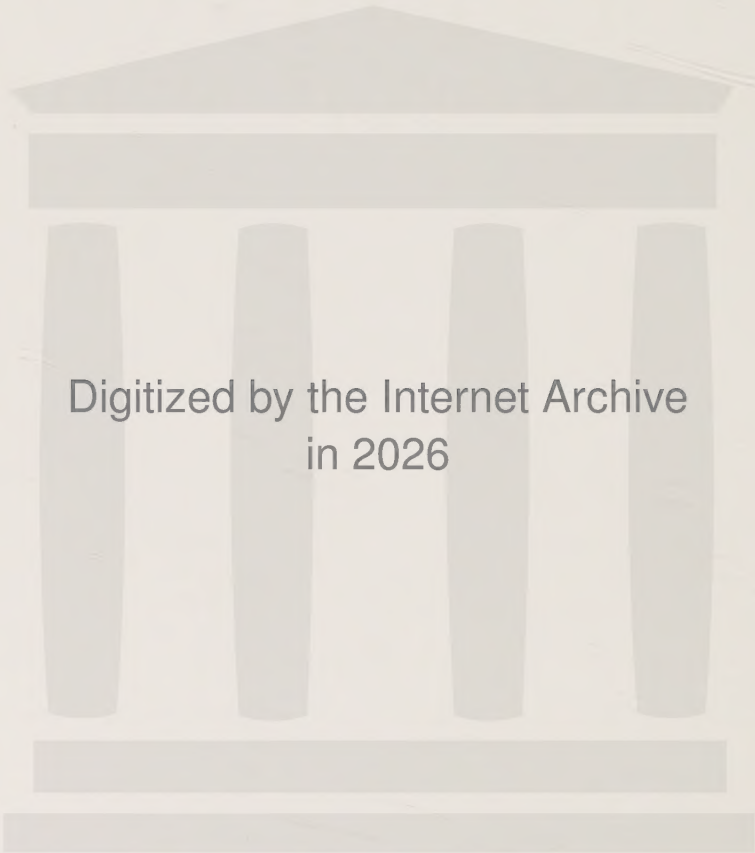
DEUXIÈME THÈSE

Propositions données par la Faculté

Procédés arithmétiques pour
fabriquer des nombres aléatoires

Soutenues le 8 Janvier 1968 devant la Commission d'examen

MM.	D. DUGUE	Président
	J.P. BENZECRI M. GIRAULT	Examinateurs
	G. MORLAT	Invité


Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41523703_0085

PUBLICATIONS DE
L'INSTITUT de
STATISTIQUE
DE L'UNIVERSITÉ DE PARIS

SOMMAIRE

INTRODUCTION 1

1. TESTS D'HYPOTHESES SIMPLES CONCERNANT PLUSIEURS PARAMETRES DANS UN PROCESSUS DE MARKOV STATIONNAIRE A TEMPS DISCRET .. 3

2. TESTS D'HYPOTHESES COMPOSEES 32

3. UNE CARACTERISATION DE LA CONVERGENCE DES PROCEDURES STATISTIQUES 60

4. ESTIMATION POUR DES PROCESSUS STOCHASTIQUES 73

5. ESTIMATION POUR DES PROCESSUS STOCHASTIQUES (CAS DE PLUSIEURS PARAMETRES) 109

BIBLIOGRAPHIE 120

VOLUME XVII FASCICULE 1 1968
PARIS

UNIV.
LEIDEN
BIBL.

Publ. Inst. Statist. Univ. Paris
17 (1968), 1-124

TABLE DES MATIERES

Pages

INTRODUCTION 1

1. TESTS D'HYPOTHÈSES SIMPLES CONCERNANT PLUSIEURS PARAMÈTRES DANS UN PROCESSUS DE MARKOV STATIONNAIRE A TEMPS DISCRET 3

1.0. Introduction 3

1.1. Préliminaires 4

1.2. Tests avec la meilleure puissance constante sur une famille de surfaces 11

1.3. Convergence des tests 18

1.4. Tests les plus stricts 20

1.5. Régions de confiance 23

1.6. Exemples 26

2. TESTS D'HYPOTHÈSES COMPOSÉES 32

2.0. Introduction 32

2.1. Cas d'un seul paramètre spécifié 33

2.2. Quelques résultats généraux 35

2.3. Tests d'hypothèses composées 42

2.3. A. Hypothèses composées linéaires 42

2.3. B. Hypothèses composées générales 54

3. UNE CARACTÉRISATION DE LA CONVERGENCE DES PROCÉDURES STATISTIQUES 60

3.0. Introduction 60

3.1. Préliminaires 61

3.2. Convergence des tests 63

3.2. A. Hypothèses simples 63

3.2. B. Discrimination 66

3.2. C. Hypothèses composées 67

3.3. Convergence des estimateurs 71

Pages

4. ESTIMATION POUR DES PROCESSUS STOCHASTIQUES 73

4.0. Introduction 73

4.1. Préliminaires 76

4.2. Définition de l'efficacité asymptotique et méthode du maximum de vraisemblance 81

4.3. Estimation pour des processus de Markov 85

4.4. Estimation pour des processus à accroissements indépendants 104

5. ESTIMATION POUR DES PROCESSUS STOCHASTIQUES (CAS DE PLUSIEURS PARAMÈTRES) 109

5.0. Introduction 109

5.1. Préliminaires 109

5.2. Définition de l'efficacité asymptotique et méthode du maximum de vraisemblance 113

5.3. Estimation pour des processus de Markov 118

BIBLIOGRAPHIE 120

INTRODUCTION

Cette étude est consacrée à des problèmes de tests d'hypothèses concernant les paramètres et d'estimation de paramètres pour les processus de Markov, strictement stationnaires à temps discret. Soit $\{X_n ; n \in \mathbf{N}^*\}$ un processus de Markov dont la probabilité de transition stationnaire est donnée par

$$P(x, A; \underset{\sim}{\theta}) = \Pr\{X_{n+1} \in A \mid X_n = x ; \underset{\sim}{\theta}\}$$

où $\underset{\sim}{\theta}$ est un paramètre vectoriel inconnu à k composantes appartenant à Θ , sous-ensemble de $\mathbf{R}^k$. Les problèmes qui se posent d'une manière assez naturelle sont de tester des hypothèses concernant les paramètres et de faire une estimation des vraies valeurs des paramètres à partir d'une réalisation finie du processus.

Dans le chapitre I, on considère le problème du test d'hypothèse simple concernant $\underset{\sim}{\theta}$ et une théorie des échantillons de grande taille est développée. On propose un critère de test semblable à celui de Rao [44] pour le cas de variables aléatoires indépendantes, et on établit ses propriétés optimales asymptotiques. Il est démontré que le test a asymptotiquement la meilleure puissance constante sur une famille de surfaces dans l'espace des paramètres, est convergent et est asymptotiquement le plus strict.

On montre également les propriétés optimales asymptotiques des régions de confiance obtenues en utilisant ce critère du test.

Le problème des tests d'hypothèses composées est traité au chapitre II. On propose un critère de test pour de telles hypothèses et ses propriétés optimales asymptotiques sont démontrées. Il est démontré, comme dans le cas de l'hypothèse simple que, pour un choix particulier des estimateurs pour les paramètres non spécifiés, le test a asymptotiquement la meilleure puissance constante sur une famille de surfaces dans l'espace des paramètres et est asymptotiquement le plus strict.

On voit aussi que, pour un choix convenable des estimateurs pour les paramètres non spécifiés, ce critère conduit pour les échantillons de grande taille, aux critères des tests utilisés couramment comme celui de Rao ([50] , p. 350), celui de Wald [54] et celui du rapport de vraisemblance. En même temps, on peut donc choisir entre plusieurs types d'estimateurs pour les paramètres non-spécifiés.

Au chapitre III, on donne une caractérisation de la convergence des tests par l'intermédiaire d'une fonction définie dans l'espace des mesures de probabilités, qui montre explicitement que, chaque fois qu'il existe un test convergent pour une hypothèse simple, le lemme de Neyman-

Pearson donne un des tests convergents les plus puissants. Le concept de convergence d'un estimateur est aussi examiné avec la même fonction.

Les chapitres IV et V traitent du problème de l'estimation. Le problème est considéré dans le cas général des processus stochastiques stationnaires et les résultats sont appliqués aux processus de Markov et aux processus à accroissements indépendants stationnaires. Dans le chapitre IV on formule essentiellement une nouvelle définition dans le cas général de l'efficacité asymptotique d'un estimateur, qui réponde à plusieurs critiques concernant le concept d'efficacité et on démontre que la méthode du maximum de vraisemblance donne des estimateurs asymptotiquement efficaces sous des hypothèses de régularité assez générales. Enfin, au chapitre V, les résultats sont étendus au cas d'un paramètre vectoriel.

Nous souhaitons exprimer notre profonde gratitude à Monsieur le Professeur D. Dugué qui nous a guidé par ses conseils très utiles tout au long de ce travail.

Monsieur le Professeur J.P. Benzécri nous a inspiré en plusieurs manières. Nous l'en remercions ainsi que d'avoir accepté de participer au jury.

Nous aimerions aussi adresser nos remerciements à Monsieur le Professeur M. Girault pour avoir accepté d'être membre du Jury.

Nous voudrions également mentionner les suggestions très utiles de Monsieur le Professeur G. Morlat à de nombreux stades de ce travail.

Monsieur le Professeur Bui Trong Lieu n'a jamais hésité à nous consacrer généreusement son temps et ses remarques nous ont conduit à de nombreuses améliorations importantes tant sur le fond que sur la présentation. Qu'il en soit remercié avec beaucoup d'appréciation.

Nous adressons aussi notre gratitude à Monsieur D. Carton qui nous a aidé de plusieurs façons.

Monsieur le Professeur C.R. Rao est la première source d'inspiration de ce travail. Nous lui exprimons notre gratitude pour ses nombreuses suggestions et pour ses encouragements.

Enfin, il nous reste le plaisir de remercier les chercheurs, les membres du Secrétariat et de la bibliothèque de l'ISUP pour diverses formes d'aide qui nous ont été apportées et pour avoir rendu agréable notre séjour à l'Institut.

1. Tests d'hypothèses simples concernant plusieurs paramètres dans un processus de Markov stationnaire à temps discret

1.0. Introduction.

Dans ce premier chapitre, le problème des tests d'hypothèses simconcernant plusieurs paramètres dans des processus de Markov à temps discret sera considéré. Plus précisement, soit $\{X_n, n \in \mathbf{N}^*\}$ un processus de Markov à temps discret, strictement stationnaire, à valeurs dans un espace des états $(\mathcal{X}, \mathcal{S})$ - espace polonais - avec la mesure de probabilité de transition stationnaire.

(1.0.1.) $P(x, A; \underset{\sim}{\theta}) = \Pr\{X_{n+1} \in A \mid X_n = x ; \underset{\sim}{\theta}\}$

pour chaque $n \in \mathbf{N}^*$ et $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)$ où $\underset{\sim}{\theta}$ est un paramètre vectoriel inconnu à k composantes apartenant à Θ , ensemble compact de $\mathbf{R}^k$ .Le problème est de tester l'hypothèse simple

(1.0.2.) $H_o : \underset{\sim}{\theta} = \underset{\sim}{\theta}_o$ où $\underset{\sim}{\theta}_o = (\theta_1^o, \ldots, \theta_k^o) \in \Theta$,
est un point intérieur de Θ .

A la différence du cas d'un seul paramètre, le graphe de la fonction de puissance d'un test sera une surface dans $\mathbf{R}^{k+1}$. Si l'on veut que le graphe de la fonction de puissance d'un test d'hypothèse H_o ait une courbure maximum le long de chaque section transversale au point $\underset{\sim}{\theta}_o$, c'est-à-dire que le test soit le plus puissant contre toutes les hypothèses simples dans un voisinage de $\underset{\sim}{\theta}_o$ soumis aux conditions de seuil, on ne peut pas le réaliser même dans la plupart des cas ordinaires. Pour étudier cette situation, Neyman et Pearson [40] ont considéré "Les régions critiques de type c" telles que la région critique soit non seulement d'un seuil donné et sans biais, mais aussi ait une puissance constante dans un voisinage infinitésimal de l'hypothèse nulle, le long d'une famille donnée d'ellipses concentriques. Wald [54] a traité le problème de plusieurs paramètres et a développé une théorie asymptotique pour les problèmes de tests d'hypothèses. On y démontre que des estimateurs du maximum de vraisemblance, racines, si elles existent, des équations de vraisemblance correspondantes, qui sont des estimateurs convergents et qui suivent asymptotiquement une loi normale, donnent des tests asymptotiquement optimaux pour de telles hypothèses contre une classe d'hypothèses alternatives simples. Rao [44] a proposé un test qui n'exige pas le calcul

des estimateurs du maximum de vraisemblance. Issacson [27] aussi a traité le même problème. Au lieu de maximiser la courbure autour de l'hypothèse nulle à travers chaque section transversale, il a proposé de maximiser la courbure totale (courbure gaussienne) au point de l'hypothèse nulle. Toutes ces études ont été faites dans le cas de variables aléatoires indépendantes, et de même loi. Dans ce chapitre, on étudie un cas plus général, le problème du test d'hypothèse simple concernant plusieurs paramètres dans un processus de Markov à temps discret, strictement stationnaire. Billingsley [13] a étudié la méthode du rapport de vraisemblance pour ce problème, mais les propriétés du test ne sont pas données. Les propriétés d'un test semblable à celui de Rao [ibid] sont présentées dans les paragraphes I.2 à I.5. La section I.1 donne des préliminaires. Dans le paragraphe I.2, il est démontré que le test que nous allons proposer a asymptotiquement la meilleure puissance constante sur une famille de surfaces. On démontre la convergence du test dans le paragraphe I.3. Dans le paragraphe suivant, I.4, on démontre que le test est asymptotiquement le plus strict. Ensuite, dans le paragraphe I.5, on démontre comment on peut se servir de ce critère de test pour construire des régions de confiance asymptotiquement optimales. Enfin, dans le paragraphe I.6 cette méthode de construction du test est illustrée par deux exemples, la loi multinomiale et la loi gamma à trois paramètres.

1.1. Préliminaires.

Soit $(\Omega, \mathcal{B}, \Pr(.\,;\underset{\sim}{\theta}))$ un espace probabilisé où la mesure de probabilité $\Pr(.\,;\underset{\sim}{\theta})$ a un paramètre inconnu $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k) \in \Theta$, où Θ est un ensemble compact de $\mathbf{R}^k$. Soit $\{X_n, n \in \mathbf{N}^*\}$ un processus de Markov à temps discret, à valeurs dans un espace des états $(\mathcal{X}, \mathcal{S})$ (espace polonais) déterminé par la probabilité de transition stationnaire

$$(1.1.1.)\quad P(x, A; \underset{\sim}{\theta}) = \Pr\{X_{n+1} \in A \mid X_n = x;\ \underset{\sim}{\theta}\}, \forall\, n \in \mathbf{N}^*, \text{ et } A \in \mathcal{S}.$$

On rappelle que $P(x, A\,;\underset{\sim}{\theta})$ est une fonction mesurable de x pour $A \in \mathcal{S}$ fixé et une probabilité sur $\mathcal{S}$ pour $x \in \mathcal{X}$ fixé. On suppose qu'il existe une probabilité stationnaire absolue $P(.\,;\theta)$ sur $\mathcal{S}$ telle que

$$(1.1.2.)\quad P(A\,;\theta) = \int_{\mathcal{X}} P(x, A\,;\underset{\sim}{\theta})\ P(dx\,;\underset{\sim}{\theta}).$$

(L'existence d'une telle mesure peut se démontrer sous certaines hypothèses sur les probabilités de transition, cf. Doob, ch. V).

Sans nuire à la généralité, on peut supposer que $P(.\ ;\underset{\sim}{\theta})$ est aussi la probabilité initiale du système. (puisque la probabilité initiale ne joue aucun rôle significatif ; pour simplifier les notations, nous avons fait cette hypothèse. D'ailleurs, ceci n'est pas anormal même si ce n'est pas nécessaire). Pour les mêmes raisons, les espérances mathématiques dans ce qui suit sont aussi prises en supposant que $P(.\ ;\underset{\sim}{\theta})$ est la probabilité initiale.

Alors, la probabilité que le système soit dans un ensemble $A \in \mathcal{S}$ après n étapes, est donnée par

$$(1.1.3.)\quad \begin{cases} P^{(1)}(.\ , A\ ;\underset{\sim}{\theta}) = \Pr(X_n \in A\ ;\underset{\sim}{\theta}) = P(A\ ;\underset{\sim}{\theta}) \text{ pour } n = 1 \\ P^{(n)}(A\ ;\underset{\sim}{\theta}) = \int_{\mathcal{X}} P^{(n-1)}(x,A;\underset{\sim}{\theta})\, P(dx;\underset{\sim}{\theta}) \text{ pour } n>1 \end{cases}$$

où $P^{(n)}(x, .\ ;\underset{\sim}{\theta})$ sont des probabilités de transition en n étapes, données par l'équation de Chapman-Kolmogorov de la façon suivante

$$(1.1.4.)\quad \begin{cases} P^{(1)}(x,A\ ;\underset{\sim}{\theta}) = P(x, A;\underset{\sim}{\theta}) \\ P^{(n+1)}(x, A;\underset{\sim}{\theta}) = \int_{\mathcal{X}} P^{(n)}(y, A;\theta)\, P(x,dy;\underset{\sim}{\theta}) \end{cases},$$

et ce sont aussi les probabilités de transition stationnaires.

Dans ce qui va suivre, les hypothèses suivantes sont faites

Hypothèse 1.1. (condition d'identifiabilité)

Les probabilités de transition, et les probabilités absolues stationnaires sont telles que

$$P(x, .\ ;\underset{\sim}{\theta}_1) = P(x, .\ ;\underset{\sim}{\theta}_2)\ \forall\ x \in \mathcal{X}$$

$$P(,\ ;\underset{\sim}{\theta}_1) = P(.\ ;\underset{\sim}{\theta}_2) \text{ si et seulement si } \underset{\sim}{\theta}_1 = \underset{\sim}{\theta}_2 .$$

Avant de faire l'hypothèse suivante, nous munissons Θ de la tribu $\mathcal{C}$, trace de la tribu borélienne de $\mathbf{R}^k$ sur Θ.

Hypothèse 1.2.

Il existe une mesure λ, σ-finie, sur $\mathcal{S}$, par rapport à laquelle toutes les probabilités de transition et les probabilités absolues stationnaires sont absolument continues,

c'est-à-dire $\forall x \in \mathcal{X}$, $P(x,. ; \underset{\sim}{\theta}) << \lambda$, $P(.; \underset{\sim}{\theta}) << \lambda$.

On a donc

$$\left.\begin{aligned} P(x,A; \underset{\sim}{\theta}) &= \int_A p(x,y; \underset{\sim}{\theta})\ \lambda(dy) \\ P(A ; \underset{\sim}{\theta}) &= \int_A p(x ; \underset{\sim}{\theta})\ \lambda(dx) \end{aligned}\right\} \forall A \in \mathcal{S} \qquad (1.1.5.)$$

On suppose que $p(x,y; \underset{\sim}{\theta})$ et $p(x ; \underset{\sim}{\theta})$ sont respectivement $\mathcal{S} \otimes \mathcal{S} \otimes \mathcal{C}$ mesurables et $\mathcal{S} \otimes \mathcal{C}$ mesurables.

Maintenant, supposons que $(X_1, \ldots, X_n)$ est un échantillon de taille n du processus défini par (1.1.1.). Un échantillon de taille n est donc un point dans $\mathcal{X}^{(n)} = \mathcal{X} \times \ldots \times \mathcal{X}$ (n fois) et noté $\underset{\sim}{X}^{(n)}$.

La fonction de vraisemblance d'un échantillon de taille n est donnée par

$$L_n(\underset{\sim}{X}^{(n)}; \underset{\sim}{\theta}) = p(X_1 ; \underset{\sim}{\theta}) \prod_{i=1}^{n-1} p(X_i, X_{i+1} ; \underset{\sim}{\theta}) \qquad (1.1.6.)$$

où $\log L_n(\underset{\sim}{X}^{(n)}; \underset{\sim}{\theta}) = \log p(X_1 ; \underset{\sim}{\theta}) + \sum_{i=1}^{n-1} \log p(X_i, X_{i+1} ; \theta)$.

Comme nous n'avons qu'une seule observation sur l'état initial, pour n assez grand, on peut, sans restreindre la généralité, prendre le terme dominant en ignorant le premier. On peut donc considérer la fonction de vraisemblance comme

$$\log L_n(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}) \approx \sum_{i=1}^{n-1} \log p(X_i, X_{i+1} ; \underset{\sim}{\theta}) \qquad (1.1.7.)$$

qui sera d'ailleurs exacte si l'on suppose que la loi initiale est concentrée en un seul point.

Hypothèse 1.3.

Pour chaque n , il existe un ensemble $D^{(n)} \in \mathcal{S}^{(n)}$ indépendant de $\underset{\sim}{\theta} \in \Theta$ tel que

$$L_n(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}) > 0 \qquad \forall\, \underset{\sim}{X}^{(n)} \in D^{(n)}$$

et $$P_n(D^{(n)} ; \underset{\sim}{\theta}) = 1 \qquad \forall\, \underset{\sim}{\theta} \in \Theta$$

où $$P_n(A^{(n)} ; \underset{\sim}{\theta}) = \int_{A_1} P(dx_1 ; \underset{\sim}{\theta}) \int_{A_2} P(x_1, dx_2 ; \underset{\sim}{\theta}) \ldots$$

$$\ldots \int_{A_n} P(x_{n-1}, dx_n ; \underset{\sim}{\theta})$$

où $A^{(n)} = A_1 \times \ldots \times A_n$.

Hypothèse 1.4.

Les dérivées partielles $\dfrac{\partial p(x,y ; \underset{\sim}{\theta})}{\partial \theta_i}$, $\dfrac{\partial^2 p(x,y ; \underset{\sim}{\theta})}{\partial \theta_i \, \partial \theta_j}$ et

$\dfrac{\partial^3 p(x,y ; \underset{\sim}{\theta})}{\partial \theta_i \, \partial \theta_j \, \partial \theta_l}$ existent pour i, j, l = 1,...,k et sont continues en $\underset{\sim}{\theta} \in \mathbf{R}^k$.

En outre, $\dfrac{\partial}{\partial \theta_i} \int_{D^{(n)}} L_n \, d\lambda^{(n)} = \int_{D^{(n)}} \dfrac{\partial}{\partial \theta_i} L_n \, d\lambda^{(n)}$

$$\frac{\partial^2}{\partial \theta_i \, \partial \theta_j} \int_{D^{(n)}} L_n \, d\lambda^{(n)} = \int_{D^{(n)}} \frac{\partial^2 L_n}{\partial \theta_i \, \partial \theta_j} \, d\lambda^{(n)} ,$$

et il existe une fonction H(x,y) , telle que

$$\frac{\partial^3}{\partial \theta_i \, \partial \theta_j \, \partial \theta_l} \log p(x,y ; \underset{\sim}{\theta}) \leqslant H(x,y) \quad \text{et} \quad E_{\underset{\sim}{\theta}}(|H(x,y)|) < \infty$$

$$\forall\, \underset{\sim}{\theta} \in \Theta \quad \text{et} \quad i, j, l = 1, \ldots, k .$$

Posons

(1.1.8.) $$E_{\underset{\sim}{\theta}}\left[\frac{\partial}{\partial \theta_i} \log p(x,y ; \underset{\sim}{\theta}) . \frac{\partial}{\partial \theta_j} \log p(x,y ; \underset{\sim}{\theta})\right] = c_{ij}(\underset{\sim}{\theta})$$

pour i, j = 1,...,k et $\Gamma(\underset{\sim}{\theta}) = \left[c_{ij}(\underset{\sim}{\theta})\right]_{k \times k}$.

Nous supposons que $\Gamma(\underset{\sim}{\theta})$ est définie positive et continue pour $\underset{\sim}{\theta} \in \Theta$. En outre, il existe deux constantes $0 < K_1$, $K_2 < \infty$, telles que

$$\underset{\underset{\sim}{\theta} \in \Theta}{\text{Inf}} \quad \text{dét } \Gamma(\underset{\sim}{\theta}) \geqslant K_1 \quad \text{et} \quad \underset{\underset{\sim}{\theta} \in \Theta}{\text{Sup}} \quad \text{dét } \Gamma(\underset{\sim}{\theta}) \leqslant K_2 \ .$$

Ceci implique aussi que les $c_{ij}(\underset{\sim}{\theta})$ sont continus pour $\underset{\sim}{\theta} \in \Theta$ et les paramètres $\theta_1, \ldots, \theta_k$ sont linéairement indépendants, c'est-à-dire que les paramètres qu'on considère ne sont pas redondants.

Précédemment, et dans ce qui suit, $E_{\underset{\sim}{\theta}}$ désigne l'espérance mathématique calculée sous l'hypothèse que $\underset{\sim}{\theta}$ est la vraie valeur du paramètre et $P(. \ ; \underset{\sim}{\theta})$ est la probabilité initiale. Cependant, cela est seulement un moyen de calculer l'espérance mathématique et la probabilité initiale ne joue aucun rôle significatif dans les calculs de l'espérance mathématique conditionnelle.

Hypothèse 1.5.

Pour chaque $\underset{\sim}{\theta} \in \Theta$, la probabilité absolue stationnaire qui existe et qui est unique par hypothèse, possède la propriété que, pour chaque $x \in \mathcal{X}$, $P(x, . \ ; \underset{\sim}{\theta}) \ll P(. \ ; \underset{\sim}{\theta})$.

Si, par exemple, l'espace des états est fini et le processus de Markov est ergodique sans sous-classe cyclique et sans la classe transitoire, l'hypothèse I.5 est satisfaisante (Billingsley [13]).

Notons par

$$\Psi(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}) = \log L_n(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta})$$

$$(1.1.9.) \qquad y_n^{(1)}(\underset{\sim}{\theta}) = \frac{1}{\sqrt{n-1}} \frac{\partial \Psi}{\partial \theta_1}$$

$$\underset{\sim}{Y}_n^T(\underset{\sim}{\theta}) = (y_n^{(1)}(\underset{\sim}{\theta}), \ldots, y_n^{(k)}(\underset{\sim}{\theta}))$$

et $$Pr(\underset{\sim}{Y}_n(\underset{\sim}{\theta}) \leqslant \underset{\sim}{t} \ ; \underset{\sim}{\theta}) = \Phi_n(\underset{\sim}{t} \ ; \underset{\sim}{\theta})$$

(Dorénavant, tous les vecteurs $(\underset{\sim}{.})$ sont des vecteurs colonnes. Leurs transposées sont notées $(\underset{\sim}{.})^T$).

PROPOSITION 1.1.1. — *Sous les hypothèses* I.1 *à* I.5 *ci-dessus, la fonction de répartition* $\Phi_n(\underset{\sim}{t}\,;\underset{\sim}{\theta})$ *converge vers la fonction de répartition de la loi gaussienne d'espérance mathématique* $\underset{\sim}{\theta}$ *et de matrice de covariance* $\Gamma(\underset{\sim}{\theta})$.

Ceci est démontré en appliquant le théorème central limite de Lévy-Lindeberg pour les martingales ([15] et [26]).

De la proposition 1.1.1., on déduit le résultat bien connu donné dans la proposition suivante

PROPOSITION 1.1.2. — *La forme quadratique*

$$(1.1.10.)\qquad Q_n(\underset{\sim}{X}^{(n)}\,;\underset{\sim}{\theta}) = \underset{\sim}{Y}_n^T(\underset{\sim}{\theta})\,\Gamma^{-1}(\underset{\sim}{\theta})\,\underset{\sim}{Y}_n(\underset{\sim}{\theta})$$

suit asymptotiquement une loi de χ^2 *avec* k *degrés de liberté.*

Remarque. — Pour les deux propositions ci-dessus, les hypothèses peuvent être affaiblies considérablement. Par exemple, si $\Gamma(\underset{\sim}{\theta})$ est de rang $r < k$, la loi asymptotique de $\underset{\sim}{Y}_n(\underset{\sim}{\theta})$ est encore gaussienne.

Par conséquent, un résultat analogue à la proposition 1.1.2. peut être obtenu comme suit.

Considérons une matrice non singulière C telle que $C^T \Gamma C = \Delta$, où Δ est une matrice diagonale d'ordre k, mais avec seulement r éléments positifs et les autres nuls. Notons par ∇ la matrice obtenue en remplaçant les éléments non nuls de Δ par leurs inverses. Or, on a $\Gamma = \Gamma C^T \nabla C \Gamma$, Aussi, la matrice de covariance asymptotique de $\nabla^{1/2} C^T \underset{\sim}{Y}_n(\underset{\sim}{\theta})$ est donnée par $\nabla^{1/2} C^T \Gamma(\theta) C \nabla^{1/2} = I_r$, où I_r est une matrice diagonale d'ordre $(k \times k)$ avec r éléments égaux à l'unité et les autres égaux à 0. Donc, on peut énoncer la

PROPOSITION 1.1.2.bis. — *La forme quadratique*

$$Q_n(\underset{\sim}{X}^{(n)}\,;\underset{\sim}{\theta}) = \underset{\sim}{Y}_n^T(\underset{\sim}{\theta})\,C\nabla\,C^T\,\underset{\sim}{Y}_n(\underset{\sim}{\theta})$$

suit asymptotiquement une loi de χ^2 *avec* r *degrés de liberté.*

LEMME 1.1.1. — *Soit* $(X_1, X_2) = \underset{\sim}{X}$ *une variable aléatoire qui suit la loi normale d'espérance mathématique* $(0,0)$ *et dont la matrice de covariance est la matrice identité, c'est-à-dire que* X_1 *et* X_2 *sont deux variables aléatoires indépendantes, chacune suivant la loi normale réduite.*

Considérons la région $W = X_1^2 + X_2^2 \leqslant d$. *Considérons aussi le cercle* $K_c : \theta_1^2 + \theta_2^2 = c$ *dans le même espace. Alors, pour tout couple*

$$\xi = (\xi_1, \xi_2) \text{ et } \eta_1, \eta_2)$$

de points appartenant à K_c

$$\int_W \exp \{ -(1/2) [(X_1 - \xi_1)^2 + (X_2 - \xi_2)^2] \} \, dX_1 \, dX_2$$

$$= \int_W \exp \{ -(1/2) [(X_1 - \eta_1)^2 + (X_2 - \eta_2)^2] \} \, dX_1 \, dX_2$$

DÉMONSTRATION . — Faisons les transformations

$$X_1 = r \cos \alpha$$
$$X_2 = r \sin \alpha$$

Alors $\xi_1 = c \cos \beta$ $\qquad \eta_1 = c \cos \delta$

$\xi_2 = c \sin \beta$ $\qquad \eta_2 = c \sin \delta$

où β et δ sont deux constantes.

De plus, $(X_1 - \xi_1)^2 + (X_2 - \xi_2)^2 = [(r \cos \alpha - c \cos \beta)^2 + (r \sin \alpha - c \sin \beta)^2]$

$= r^2 + c^2 - 2rc \cos (\alpha - \beta)$.

Donc, $\int_W \exp \{ -(1/2) [(X_1 - \xi_1)^2 + (X_2 - \xi_2)^2] \} \, dX_1 \, dX_2$

$$= \int_0^d \int_0^{2\pi} \exp \{ -(1/2) [r^2 + c^2 - 2rc \cos (\alpha - \beta)] \} \, r \, dr \, d\alpha$$

$$= \int_0^d \exp [-(1/2)(r^2 + c^2)] \, r \, dr \int_0^{2\pi} \exp [rc \cos (\alpha - \beta)] \, d\alpha .$$

Considérons $\int_{o}^{2\pi} e^{rc \cos(\alpha - \beta)} d\alpha = f(r, c, \beta)$.

Montrons que $f(r, c, \beta)$ est indépendant de β , ce qui achève la démonstration.

$$\frac{\partial f}{\partial \beta} = \int_{o}^{2\pi} r\, c \sin(\alpha - \beta)\, e^{rc \cos(\alpha - \beta)}\, d\alpha = [e^{rc\cos(\alpha - \beta)}]_{o}^{2\pi} = 0 ;$$

donc $f(r, c, \beta)$ est indépendant de β .

Nous avons le corollaire suivant

Soit $\underset{\sim}{X}^T = (X_1, \ldots, X_k)$ *un vecteur aléatoire qui suit une loi normale à* k *dimensions d'espérance mathématique* $\underset{\sim}{\theta}_o$ *et de matrice de covariance* Λ *, définie positive.*

Soit $W = (\underset{\sim}{X} - \underset{\sim}{\theta}_o)^T \Lambda^{-1} (\underset{\sim}{X} - \underset{\sim}{\theta}_o) \leqslant d$

et $K_c = (\underset{\sim}{\theta} - \underset{\sim}{\theta}_o)^T \Lambda^{-1} (\underset{\sim}{\theta} - \underset{\sim}{\theta}_o) = c$.

Alors, pour tout couple $\underset{\sim}{\theta}_1$ et $\underset{\sim}{\theta}_2 \in K_c$,

$$P(W ; \underset{\sim}{\theta}_1) = P(W ; \underset{\sim}{\theta}_2) .$$

En effet, ce corollaire peut se déduire aussi du résultat que $(\underset{\sim}{X} - \underset{\sim}{\theta}_o)^T \Lambda^{-1} (\underset{\sim}{X} - \underset{\sim}{\theta}_o)$ suit une loi de χ^2 décentrée avec le paramètre de décentrement $(\underset{\sim}{\theta} - \underset{\sim}{\theta}_o) \Lambda^{-1} (\underset{\sim}{\theta} - \underset{\sim}{\theta}_o)$ sous l'hypothèse que $\underset{\sim}{\theta}$ est l'espérance mathématique de $\underset{\sim}{X}$.

1.2. Tests pour les hypothèses simples avec la meilleure puissance constante sur une famille de surfaces.

DÉFINITION 1.2.1. – Une fonction de test (on peut dire simplement un test) $\varphi_n(X^{(n)})$ pour tester une hypothèse simple $H_o : \underset{\sim}{\theta} = \underset{\sim}{\theta}_o$ (d'après la notation de Lehmann [33]) est dite avoir la puissance constante uniformément la meilleure sur une famille de surfaces K_c définie dans l'espace des paramètres, si les conditions suivantes sont remplies

(1) $E_{\underset{\sim}{\theta}_o}(\varphi_n(\underset{\sim}{X}^{(n)})) = \alpha \ (0 < \alpha < 1)$ où α, donné, est appelé seuil du test.

(2) $E_{\underset{\sim}{\theta}_1}(\varphi_n(\underset{\sim}{X}^{(n)})) = E_{\underset{\sim}{\theta}_2}(\varphi_n(\underset{\sim}{X}^{(n)}))$ pour tout couple $\underset{\sim}{\theta}_1$ et $\underset{\sim}{\theta}_2$ appartenant à la même surface K_c .

(3) Quel que soit $\varphi_n^*(\underset{\sim}{X}^{(n)})$, un autre test qui satisfait aux conditions (1) et (2)

$$E_{\underset{\sim}{\theta}}(\varphi_n^*) \leqslant E_{\underset{\sim}{\theta}}(\varphi_n) \quad , \forall \ \underset{\sim}{\theta} \in K_c \ .$$

Cette définition est essentiellement due à Wald [54] , sauf qu'ici, la classe des tests est élargie pour inclure aussi des tests randomisés.

PROPOSITION 1.2.1. — *Soit* $\underset{\sim}{Z}^T = (Z_1,..., Z_k)$ *un vecteur aléatoire qui suit une loi normale à* k *dimensions avec le vecteur des espérances mathématiques* $\underset{\sim}{\theta}^T = (\theta_1,..., \theta_k)$ *inconnu et la matrice de covariance* $\Lambda = [\sigma_{ij}]$, *définie positive, connue. Soit* $\underset{\sim}{Z}$ *une observation dont la densité est notée* $L(\underset{\sim}{Z} ; \underset{\sim}{\theta})$.

Pour tester l'hypothèse $H_o : \underset{\sim}{\theta} = \underset{\sim}{\theta}_o = (\theta_1^o,..., \theta_k^o)$, *considérons la fonction de test définie par*

$$\varphi(\underset{\sim}{Z}) \begin{cases} = 1 \ si \ \underset{\sim}{Z} \in W \\ = 0 \ sinon, \end{cases}$$

où $$W = \left\{ \underset{\sim}{Z} \mid \sum_{i,j=1}^{k} \sigma_{ij} \, y^{(i)}(\underset{\sim}{\theta}_o) \, y^{(j)}(\underset{\sim}{\theta}_o) \geqslant d \right\}$$

$$y^{(i)}(\underset{\sim}{\theta}_o) = \left(\frac{\partial \log L}{\partial \theta_i} \right)_{\underset{\sim}{\theta} = \underset{\sim}{\theta}_o} \quad \text{pour } i = 1,..., k$$

et d *est choisi pour satisfaire à la condition de seuil du test.*

Alors $\varphi(\underset{\sim}{Z})$ *a la puissance constante uniformément la meilleure parmi la classe des tests invariants sous un groupe de transformations linéaires non-singulières, sur les surfaces* K_c *définies par*

$$K_c(\underset{\sim}{\theta}_o) = \sum_{i,j=1}^{k} \sigma^{ij} (\theta_i - \theta_i^o) (\theta_j - \theta_j^o) = c$$

où $[\sigma^{ij}] = [\sigma_{ij}]^{-1}$.

DÉMONSTRATION . – La densité s'écrit

$$L(\underset{\sim}{Z} ; \underset{\sim}{\theta}) = \frac{1}{(2\pi)^{k/2}\sqrt{\det \Lambda}} \exp \{ -\frac{1}{2} (\underset{\sim}{Z} - \underset{\sim}{\theta})^T \Lambda^{-1} (\underset{\sim}{Z} - \underset{\sim}{\theta}) \}$$

Donc $$\frac{\partial \text{Log } L}{\partial \theta_i} = \sum_{j=1}^{k} \sigma^{ij} (Z_j - \theta_j) \quad \text{pour } i = 1,\ldots, k$$

où $$\underset{\sim}{Y}^T(\underset{\sim}{\theta}) = (y^{(1)}(\underset{\sim}{\theta}),\ldots, y^{(k)}(\underset{\sim}{\theta})) = \left(\frac{\partial \text{Log } L}{\partial \theta_1}, \ldots, \frac{\partial \text{Log } L}{\partial \theta_k} \right)$$

$$= (\underset{\sim}{Z} - \underset{\sim}{\theta})^T \Lambda^{-1} \ .$$

La région W définie dans la proposition est donnée par

$$W = \{ \underset{\sim}{Z} \mid \underset{\sim}{Y}^T (\underset{\sim}{\theta}_o) \Lambda \underset{\sim}{Y}(\underset{\sim}{\theta}_o) \geqslant d \}$$

$$= \{ \underset{\sim}{Z} \mid (\underset{\sim}{Z} - \underset{\sim}{\theta}_o)^T \Lambda^{-1} (\underset{\sim}{Z} - \underset{\sim}{\theta}_o) \geqslant d \} \ .$$

Puisque $\underset{\sim}{Z}$ suit une loi normale de vecteur d'espérance mathématique $\underset{\sim}{\theta}_o$ et de matrice de covariance Λ sous H_o , la condition (2) est satisfaite par le lemme 1.1.1. Pour démontrer la condition (3) nous utiliserons le résultat de Wald [54] selon lequel φ est un des tests les plus puissants pour H_o contre une alternative convenablement choisie.

Considérons des transformations non singulières dans l'espace des paramètres données par

(1.2.4.) $$\theta'_i - \theta^\circ_i = \beta_{i1}(\theta_1 - \theta_1^\circ) + \ldots + \beta_{ik}(\theta_k - \theta_k^\circ)$$

pour $i = 1,\ldots,k$ qui transforment la surface K_c en la sphère,

(1.2.5.) $$K'_c : (\underset{\sim}{\theta}' - \underset{\sim}{\theta}_o)^T (\underset{\sim}{\theta}' - \underset{\sim}{\theta}_o) = c \ .$$

Considérons des transformations dans l'espace des observations

(1.2.6.) $$Z'_i - \theta^\circ_i = \beta_{i1} (Z_1 - \theta_1^\circ) + \ldots + \beta_{ik}(Z_k - \theta_k^\circ)$$

pour $i = 1,\ldots k$. Alors les variables $Z'_1,\ldots,Z'_k$ sont indépendantes et chaque Z'_i suit la loi normale d'espérance mathématique θ_i° et de variance unité sous H_o . La région W se transforme en

(1.2.7.) $\{ \underset{\sim}{Z}' \mid (\underset{\sim}{Z}' - \underset{\sim}{\theta}_o)^T (\underset{\sim}{Z}' - \underset{\sim}{\theta}_o) \geqslant d \}$.

Soit φ^* un autre test invariant qui satisfait aux conditions (1) et (2). Soient $\varphi^{*\prime}$ et φ' les images de φ^* et φ par les transformations ci-dessus.

Il faut démontrer que

$$E_{\underset{\sim}{\theta}} (\varphi) \geqslant E_{\underset{\sim}{\theta}} (\varphi^*) \quad \text{pour tout } \underset{\sim}{\theta} \in K_c$$

c'est-à-dire $E_{\underset{\sim}{\theta}'} (\varphi') \geqslant E_{\underset{\sim}{\theta}'} (\varphi^{*\prime})$ pour tout $\underset{\sim}{\theta}' \in K'_c$.

Notons par $P'(\,.\,; \underset{\sim}{\theta}')$, la mesure induite par la transformation et $L'(\underset{\sim}{Z}' ; \underset{\sim}{\theta}')$ sa densité.

Considérons la probabilité

(1.2.8.) $$\int_{K'_c} P'(.\,; \underset{\sim}{\theta}')\, dA$$

dont la densité est donnée par $\int_{K'_c} L'(\underset{\sim}{Z}' ; \underset{\sim}{\theta}')\, dA$.

Wald a démontré que W' est la région critique la plus puissante pour H_o contre l'hypothèse H_1 ; la probabilité est

$$\int_{K'_c} P'(.\,; \underset{\sim}{\theta}')\, dA \ .$$

Donc, pour tout test $\varphi^{*\prime}$ qui satisfait à la condition de seuil (condition 1), on a

$$\int_{\mathcal{X}} \int_{K'_c} \varphi' L'(\underset{\sim}{Z}' ; \underset{\sim}{\theta}')\, dA\, d\underset{\sim}{Z}' \geqslant \int_{\mathcal{X}} \int_{K'_c} \varphi^{*\prime} L'(\underset{\sim}{Z}' ; \underset{\sim}{\theta}')\, dA\, dZ'$$

Puisque L' est continue en $\underset{\sim}{\theta}'$ et mesurable en $\underset{\sim}{Z}'$, elle est mesurable en couple et donc on peut changer l'ordre d'intégration. Alors,

$$\int_{K'_c} E_{\underset{\sim}{\theta}'} (\varphi')\, dA \geqslant \int_{K'_c} E_{\underset{\sim}{\theta}'} (\varphi^{*\prime})\, dA \ .$$

Mais $E_{\underset{\sim}{\theta}'}(\varphi')$ est constant sur K'_c . Si φ^* aussi satisfait à la condition (2), alors $E_{\underset{\sim}{\theta}'}(\varphi^{*\prime})$ est constant sur K'_c et l'on a

$$E_{\underset{\sim}{\theta}'}(\varphi') \geqslant E_{\underset{\sim}{\theta}'}(\varphi^{*\prime}) , \quad \forall \underset{\sim}{\theta}' \in K'_c .$$

Puisque les transformations linéaires que nous avons effectuées sont non singulières, il en découle que

$$E_{\underset{\sim}{\theta}}(\varphi) \geqslant E_{\underset{\sim}{\theta}}(\varphi^*) , \quad \forall \underset{\sim}{\theta} \in K_c$$

ce qui démontre la condition (3) .

Le corollaire suivant n'est qu'une manière d'exprimer la proposition précédente.

COROLLAIRE 1.2.1. — *Dans le cas d'un processus stochastique à temps discret, si la loi jointe des* k *premières variables est une loi normale à* k *dimensions avec le vecteur à espérance mathématique* $\underset{\sim}{\theta}^T = (\theta_1,\ldots,\theta_k)$ *inconnu et la matrice de covariance* Λ *définie positive connue, alors pour tester l'hypothèse* $H_o : \underset{\sim}{\theta} = \underset{\sim}{\theta}_o$ *à partir d'une seule suite d'observations* $\underset{\sim}{X}^{(k)} = (X_1,\ldots, X_k)$, *considérons la région critique*

$$(1.2.11) \quad W : \{ \underset{\sim}{X}^{(k)} \mid Y^T(\underset{\sim}{\theta}_o) \Lambda \underset{\sim}{Y}(\underset{\sim}{\theta}_o) \geqslant d \}$$

où $\underset{\sim}{Y}^T(\underset{\sim}{\theta}_o) = (y^{(1)}(\underset{\sim}{\theta}_o),\ldots, y^{(k)}(\underset{\sim}{\theta}_o))$.

$y^{(i)}(\underset{\sim}{\theta}_o) = \left(\dfrac{\partial \operatorname{Log} L}{\partial \theta_i}\right)_{\underset{\sim}{\theta} = \underset{\sim}{\theta}_o}$, L *étant la densité de la loi normale à* k *dimensions d'espérance mathématique* $\underset{\sim}{\theta}$ *et de matrice de covariance* Λ , d *étant choisi pour satisfaire à la condition de seuil. Cette région critique a la puissance constante uniformément la meilleure sur les surfaces*

$$(1.2.12) \qquad K_c : (\underset{\sim}{\theta} - \underset{\sim}{\theta}_o)^T \Lambda^{-1} (\underset{\sim}{\theta} - \underset{\sim}{\theta}_o) = c .$$

Dans le cas général, lorsque les lois ne sont pas gaussiennes, nous ne considérons que les tests non-randomisés, c'est-à-dire les régions critiques. De plus nous supposons que Θ est contenu dans un voisinage de $\underset{\sim}{\theta}_o$.

DÉFINITION 1.2.2. (Wald). – Une suite de régions critiques $\{W_n\}_{n \in \mathbf{N}^*}$ pour tester l'hypothèse simple $H_o : \underset{\sim}{\theta} = \underset{\sim}{\theta}_o$ est dite avoir asymptotiquement la meilleure puissance constante, sous la condition de seuil, sur les surfaces K_c dans l'espace de paramètres, si les conditions suivantes sont réalisées

(1) $\quad P_n(W_n ; \underset{\sim}{\theta}_o) = \alpha \ , \ \forall n \in \mathbf{N}^*$ (condition de seuil)

(2) $$\lim_{n \to \infty} \left\{ \sup_{c} \left[\sup_{\theta \in K_c} P_n(W_n ; \underset{\sim}{\theta}) - \inf_{\underset{\sim}{\theta} \in K_c} P_n(W_n ; \underset{\sim}{\theta}) \right] \right\} = 0 \ .$$

(3) Pour chaque suite $\{W^*_n\}_{n \in \mathbf{N}^*}$ de régions critiques qui satisfait aux conditions (1) et (2) ,

$$\lim_{n \to \infty} \left\{ \sup_{\underset{\sim}{\theta} \in \Theta} \left[P_n(W_n ; \underset{\sim}{\theta}) - P_n(W^*_n ; \underset{\sim}{\theta}) \right] \right\} = 0 \ .$$

Avant d'énoncer la proposition suivante, nous faisons l'hypothèse supplémentaire 1.6. que la convergence vers la loi normale dans la proposition 1.1.1. est uniforme en $\underset{\sim}{\theta} \in \Theta$.

La remarque suivant est évidente

REMARQUE 1.2.1. – On peut considérer $\{\underset{\sim}{Y}_n(\underset{\sim}{\theta})\}_{n \in \mathbf{N}^*}$ du paragraphe I.1 comme une suite de fonctions telles que $\forall n \in \mathbf{N}^*$, $\underset{\sim}{Y}_n$ soit une application mesurable de $(\mathcal{X}^{(n)}, \mathcal{S}^{(n)})$ dans $(\mathbf{R}^{(k)}, \mathcal{B}^{(k)})$. Pour $U^{(k)} \in \mathcal{B}^{(k)}$, posons

$$W_n = \underset{\sim}{Y}^{-1}_n (U^{(k)}) \in \mathcal{S}^{(n)} \ .$$

Alors $P_n(W_n) = P_n(\underset{\sim}{Y}^{-1}_n (U^{(k)})) = P_{\underset{\sim}{Y}_n} (U^{(k)})$.

(On supprime l'indice $\underset{\sim}{\theta}$ puisqu'il ne joue aucun rôle dans cette remarque).

La suite $\{\underset{\sim}{Y}_n\}_{n \in \mathbf{N}^*}$ converge en loi vers une variable aléatoire normale Y , c'est-à-dire

$$P_{\underset{\sim}{Y}_n} (U^{(k)}) \longrightarrow \Phi_{\underset{\sim}{Y}} (U^{(k)}) \ , \text{ où } \Phi_{\underset{\sim}{Y}} \text{ est la loi de } \underset{\sim}{Y} \ .$$

Par conséquent

$$\lim_{n \to \infty} [P_n(W_n) - \Phi_{\underset{\sim}{Y}}(U^{(k)})] = 0 \tag{1.2.13}$$

$\forall\, U^{(k)} \in \mathcal{B}^{(k)}$. Donc, on observe que les régions critiques obtenues au moyen de $\underset{\sim}{Y}_n$, peuvent toujours se tracer dans l'espace des échantillons.

Dans la proposition suivante, on considère le cas décrit dans le paragraphe 1.1.

PROPOSITION 1.2.2. – *Sous les hypothèses* 1.1. *à* 1.6 *pour tester l'hypothèse simple* $\underset{\sim}{\theta} = \underset{\sim}{\theta}_o$, *considérons pour chaque* n *la région critique* W_n *définie ci-dessous.*

Soit

$$U_n^{(k)} = \left\{ \underset{\sim}{Y}_n(\underset{\sim}{\theta}_o) \in \mathbf{R}^{(k)} \;\middle|\; \sum_{i,j=1}^{k} c^{ij}(\underset{\sim}{\theta}_o)\, y_n^{(i)}(\underset{\sim}{\theta}_o)\, y_n^{j}(\underset{\sim}{\theta}_o) \geqslant d_n \right\}$$

où $[c^{ij}(\underset{\sim}{\theta}_o)] = [c_{ij}(\underset{\sim}{\theta}_o)]^{-1}$ *et* d_n *est choisi tel que la région ait un seuil donné*

$$W_n = \underset{\sim}{Y}_n^{-1}(\underset{\sim}{\theta}_o) \circ U_n^{(k)} \quad . \tag{1.2.14}$$

Considérons la famille des surfaces K_{c_n} *dans l'espace des paramètres définie par*

$$K_{c_n} = \sum_{i,j=1}^{k} c_{ij}(\underset{\sim}{\theta}_o)\,(\theta_i - \theta_i^\circ)(\theta_j - \theta_j^\circ) = c_n \tag{1.2.15}$$

où $c_n \to 0$ *et* $n\, c_n \to \infty$ quand $n \to \infty$.

Alors $\{W_n\}_{n \in \mathbf{N}^*}$ *a asymptotiquement la meilleure puissance constante sur* K_{c_n} .

DÉMONSTRATION. – La condition (1) est vérifiée par la construction de W_n . D'après la proposition 1.1.1. avec l'hypothèse de convergence uniforme et la remarque 1.2.1.,la condition (2) est satisfaite. Enfin pour

le cas normal, la condition (3) est satisfaite par la proposition 1.2.1. et le corollaire du lemme 1.1.1. Donc la proposition est démontrée.

1.3. Convergence des tests.

Considérons d'abord le cas d'un paramètre, c'est-à-dire $k = 1$, dans le paragraphe 1.1. On sait que pour tester l'hypothèse $H_o : \theta = \theta_o$ la région critique définie par

$$(1.3.1.) \quad W_n : \left\{ X^{(n)} \;\middle|\; \left| \left(\frac{\partial \text{Log } L_n}{\partial \theta} \right) \theta = \theta_o \right| \geqslant d_n \right\}$$

où d_n est convenablement choisi, est localement la plus puissante et sans biais. Par un raisonnement assez simple, on peut démontrer que le test est aussi convergent contre toutes les hypothèses alternatives simples $\theta = \theta_1$ $(\theta_1 \neq \theta_o)$. Nous supposons que

$$(1.3.2.) \quad E_{\theta_1} \left\{ \frac{\partial^2 \log p(X_1, X_2 ; \theta}{\partial \theta^2} \right\} \text{ existe et qu'elle est bornée pour tout couple } (\theta, \theta_1) \in \Theta \times \Theta .$$

Nous savons que $P_n(W_n ; \theta_o) = \alpha$ et il faut démontrer que $P_n(W_n ; \theta_1) \longrightarrow 1$ lorsque $n \to \infty$, pour chaque $\theta_1 \neq \theta_o$.
En notant

$$(1.3.3.) \qquad Y_n(\theta) = \frac{1}{\sqrt{n-1}} \frac{\partial \text{Log } L_n}{\partial \theta}$$

$Y_n(\theta_1) = Y_n(\theta_o) + (\theta_1 - \theta_o) Y'_n(\theta^*)$ où $Y'_n(\theta^*)$ est la dérivée de $Y(\theta)$ par rapport à θ prise à la valeur θ^* où $|\theta^* - \theta_o| < |\theta_1 - \theta_o|$.
Donc

$$E_{\theta_1}(Y_n(\theta_1)) = 0 = E_{\theta_1}(Y_n(\theta_o)) + (\theta_1 - \theta_o)\sqrt{n-1}\, E_{\theta_1}\left(\frac{1}{n-1} \frac{\partial^2 \text{Log } L_n}{\partial \theta^2} \right) \theta = \theta^*$$

ce qui démontre que

$$| E_{\theta_1}(Y_n(\theta_o)) | \longrightarrow \infty \quad \text{quand } n \longrightarrow \infty .$$

Puisque $$E_\theta \left\{ \left(\frac{\partial \text{Log } p(X_1, X_2 ; \theta)}{\partial \theta} \right)^2 \right\} < \infty, \quad \forall \theta$$

il s'ensuit que $P_n (| Y_n(\theta_o) | \geqslant \delta ; \theta) \to 1$ quand $n \to \infty$, quel que soit $\delta > 0$. D'où la

PROPOSITION 1.3.1. – $\{ \underset{\sim}{X}^{(n)} \mid | Y_n(\theta_o) | \geqslant d_n \}$ *est un test pour* H_o *localement le plus puissant sans biais et convergent contre toutes les hypothèses alternatives* $\theta = \theta_1 \ (\theta_1 \neq \theta_o)$.

Remarque. – Si l'on ne fait pas l'hypothèse supplémentaire ci-dessus, mais si l'on suppose que, pour chaque suite $\{\theta_n\}$ telle que $\theta_n \to \theta_o$, on a

$$E_{\theta_n} \left\{ \left(\frac{\partial^2 \text{Log } p(X_1, X_2 ; \theta)}{\partial \theta^2} \right) \theta = \theta^{*n} \right\} \to E_{\theta_o} \left\{ \left(\frac{\partial^2 \text{Log } p(X_1, X_2 ; \theta)}{\partial \theta^2} \right) \theta = \theta_o \right\}$$

où $| \theta^{*n} - \theta_o | < | \theta_n - \theta_o |$, $\forall n \in \mathbf{N}^*$, alors le test est convergent pour toute suite $\{\theta_n\}$ telle que $\theta_n \to \theta_o$ et $\sqrt{n} \ | \theta_n - \theta_o | \to \infty$ quand $n \to \infty$. On aurait bien eu des raisons de considérer des hypothèses alternatives se rapprochant de l'hypothèse nulle lorsque n devient grand. Ce cas n'est pas inconnu et est souvent traité par plusieurs auteurs.

Considérons maintenant le cas de plusieurs paramètres, c'est-à-dire $k > 1$.

PROPOSITION 1.3.2. – *Le test défini dans la proposition* 1.2.2. *est convergent pour* H_o *contre toutes les hypothèses alternatives simples* $\theta \in K_{c_n}$.

DÉMONSTRATION. – La région critique est donnée par

$$W_n : \left\{ \underset{\sim}{X}^{(n)} \ \middle| \ Q(\underset{\sim}{X}^{(n)}) \geqslant d_n \right\}$$

où $Q(\underset{\sim}{X}^{(n)}) = \underset{\sim}{Y}_n^T (\underset{\sim}{\theta}_o) \ \Gamma^{-1} (\underset{\sim}{\theta}_o) \ \underset{\sim}{Y}_n (\underset{\sim}{\theta}_o)$.

Pour n assez grand

$$Q(\underset{\sim}{X}^{(n)}) \approx (n-1) (\underset{\sim}{\hat{\theta}}_n - \underset{\sim}{\theta}_o)^T \ \Gamma (\underset{\sim}{\theta}_o) (\underset{\sim}{\hat{\theta}}_n - \underset{\sim}{\theta}_o) ,$$

où $\hat{\underset{\sim}{\theta}}_n$ sont des estimateurs du maximum de vraisemblance (racines convergentes vu les hypothèses). $Q(\underset{\sim}{X}^{(n)})$ suit asymptotiquement une loi de χ^2 avec k degrés de liberté sous H_o et suit une loi de χ^2 décentrée avec le paramètre de décentrement $(n-1)c_n$ sous H_1. D'où il résulte que le test est convergent pour les alternatives simples $\underset{\sim}{\theta} \in K_{c_n}$.

1.4. Tests les plus stricts pour des hypothèses simples.

Soient $\underset{\sim}{\theta}_o$ et $\underset{\sim}{\theta}_1$ deux points de l'espace des paramètres. Soit $E_{P_n}(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_o ; \alpha)$, la borne supérieure de $E_{P_n}(\varphi_n ; \underset{\sim}{\theta}_1)$ par rapport à toutes les fonctions de tests φ_n pour $H_o : \underset{\sim}{\theta} = \underset{\sim}{\theta}_o$, telles que $E_{P_n}(\varphi_n \mid \underset{\sim}{\theta}_o) = \alpha$, où $0 < \alpha < 1$ est donné. Alors la puissance d'un test φ_n quelconque de seuil α basé sur un échantillon de taille n pour $H_o : \underset{\sim}{\theta} = \underset{\sim}{\theta}_o$ contre $H_1 : \underset{\sim}{\theta} = \underset{\sim}{\theta}_1$, ne peut jamais dépasser $E_{P_n}(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_o ; \alpha)$.

DÉFINITION 1.4.1. — n étant fixé, un test $\varphi_n(X^{(n)})$ est dit test le plus strict pour tester $H_o : \underset{\sim}{\theta} = \underset{\sim}{\theta}_o$ de seuil α, si

(1) $$E_{P_n}(\varphi_n \mid \theta_o) = \alpha$$

(2) $$\sup_{\underset{\sim}{\theta} \in \Theta} [E_{P_n}(\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha) - E_{P_n}(\varphi_n \mid \theta)] \leqslant \sup_{\underset{\sim}{\theta} \in \Theta} [E_{P_n}(\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha) - E_{P_n}(\varphi_n^* \mid \underset{\sim}{\theta})]$$

quel que soit le test $\varphi_n^*(\underset{\sim}{X}^{(n)})$ satisfaisant à la condition (1).

NOTA. — $E_{P_n}(. \mid \underset{\sim}{\theta})$ représente l'espérance mathématique avec la mesure de probabilité $P_n(. ; \underset{\sim}{\theta})$.

PROPOSITION 1.4.1. — *Plaçons-nous dans le cas de la proposition* 1.2.1. *Le test* $\varphi(Z)$ *qui y est défini est le plus strict pour* H_o *dans la classe des tests invariants* y *considérée.*

DÉMONSTRATION. – Raisonnons par l'absurde. Supposons que le test φ ne soit pas le plus strict pour H_o . Alors, il existe un test $\varphi^*(Z)$ tel que

$$E\psi(\varphi^* \mid \underset{\sim}{\theta}_o) = \alpha \quad \text{et}$$

$$(1.4.1.) \qquad \sup_{\underset{\sim}{\theta} \in \Theta} [E\psi(\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha) - E\psi(\varphi \mid \underset{\sim}{\theta})] > \sup_{\theta \in \Theta} [E\psi(\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha) - E\psi(\varphi^* \mid \underset{\sim}{\theta})] .$$

Ici ψ est la mesure gaussienne considérée).

Comme dans le cas de la proposition 1.2.1. ,

$E\psi(\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha)$, et $E\psi(\varphi ; \underset{\sim}{\theta})$ sont constantes sur les surfaces

$$K_c : \left\{ \sum_{i,j=1}^{k} \sigma^{ij} (\theta_i - \theta_i^\circ)(\theta_j - \theta_j^\circ) = c \right\} .$$

Donc, d'après (1.4.1.), il résulte qu'il existe une valeur c_o telle que, pour tout $\underset{\sim}{\theta} \in K_{c_o}$

$$(1.4.2.) \qquad E\psi(\varphi^* \mid \underset{\sim}{\theta}) > E\psi(\varphi \mid \underset{\sim}{\theta}) ,$$

ce qui est en contradiction avec la proposition 1.2.1. C.Q.F.D.

Dans le cas général comme au paragraphe 1.2., nous ne considérons que les tests non randomisés et donc dans la proposition suivante, φ_n est la fonction indicatrice d'une région critique. En outre, nous supposons que l'hypothèse 1.6. est aussi satisfaite, c'est-à-dire la convergence vers la loi normale dans la proposition 1.1.1. est uniforme en $\underset{\sim}{\theta} \in \Theta$.

DÉFINITION 1.4.2. (Wald). – Une suite de tests $\{\varphi_n(\underset{\sim}{X}^{(n)})\}_{n \in \mathbf{N}^*}$ pour $H_o : \underset{\sim}{\theta} = \underset{\sim}{\theta}_o$ est dite asymptotiquement la plus stricte, de seuil α, si les conditions suivantes sont réalisées

(1) $E_{P_n}(\varphi_n(\underset{\sim}{X}^{(n)}) \mid \underset{\sim}{\theta}_o) = \alpha$, $\forall n \in \mathbf{N}^*$

(2) Pour chaque suite $\{\varphi_n^*\}_{n \in \mathbf{N}^*}$ de tests de même seuil, on a

$$\limsup_{n \to \infty} \left\{ \sup_{\underset{\sim}{\theta} \in \Theta} \left[E_{P_n} (\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha) - E_{P_n} (\varphi_n \mid \underset{\sim}{\theta}) \right] - \sup_{\underset{\sim}{\theta} \in \Theta} \left[E_{P_n} (\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha) - E_{P_n} (\varphi_n^* \mid \underset{\sim}{\theta}) \right] \right\} \leqslant 0 .$$

PROPOSITION **1.4.2.** — *La suite* $\{\varphi_n\}_{n \in \mathbf{N}^*}$ *de tests définie dans la proposition* 1.2.2. *est un des tests asymptotiquement les plus stricts de seuil* α *pour* $H_o : \underset{\sim}{\theta} = \underset{\sim}{\theta}_o$ *donné par* (1.0.2.) .

DÉMONSTRATION. — Notons par $\psi(. ; \underset{\sim}{\theta})$ la loi normale à k dimensions d'espérance mathématique $\underset{\sim}{\theta}$ et de matrice de covariance $\Gamma(\underset{\sim}{\theta}_o)$. Soit $E\psi(\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha)$ la borne supérieure de $E\psi(\varphi \mid \underset{\sim}{\theta})$ par rapport à φ où φ est la fonction indicatrice d'un ensemble $U^{(k)} \in \mathcal{B}^{(k)}$ tel que $E\psi(\varphi \mid \underset{\sim}{\theta}_o) = \alpha$ [$E\psi(. \mid \underset{\sim}{\theta})$ est toujours l'espérance mathématique calculée sous l'hypothèse que $\psi(. ; \underset{\sim}{\theta})$ est la loi avec le paramètre $\underset{\sim}{\theta}$] .

D'après la proposition 1.1.1. avec la condition de convergence uniforme et d'après la remarque 1.2.1. il s'ensuit que

$$\lim_{n \to \infty} \left[E_{P_n} (\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha) - E\psi(\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha) \right] = 0 \qquad (1.4.3.)$$

uniformément en $\underset{\sim}{\theta} \in \Theta$.

Maintenant, considérons

$$K_c : \left\{ \underset{\sim}{\theta} \;\middle|\; \sum_{i,j=1}^{k} c^{ij} (\underset{\sim}{\theta}_o)(\theta_i - \theta_i^\circ)(\theta_j - \theta_j^\circ) = c \right\}$$

$E\psi(\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha)$ est constant pour les points $\underset{\sim}{\theta} \in K_c$. Donc, d'après (1.4.3.) on a

$$\lim_{n \to \infty} \left\{ \sup_{\underset{\sim}{\theta} \in K_c} E_{P_n} (\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha) - \inf_{\underset{\sim}{\theta} \in K_c} E_{P_n} (\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha) \right\} = 0. \qquad (1.4.4.)$$

Supposons que la proposition ne soit pas vraie. Alors, il existe une suite de tests $\{\varphi_n^*\}_{n \in \mathbf{N}^*}$ telle que

$$E_{P_n}(\varphi^*_n \mid \underset{\sim}{\theta}_o) = \alpha \quad \forall\, n \in \mathbf{N}^*$$

et

$$(1.4.5.) \quad \limsup_{n \to \infty} \left\{ \sup_{\underset{\sim}{\theta} \in \Theta} \left[E_{P_n}(\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha) - E_{P_n}(\varphi_n \mid \underset{\sim}{\theta}) \right] - \sup_{\underset{\sim}{\theta} \in \Theta} \left[E_{P_n}(\underset{\sim}{\theta}, \underset{\sim}{\theta}_o ; \alpha) - E_{P_n}(\varphi^*_n \mid \underset{\sim}{\theta}) \right] \right\} = \delta \ .$$

Mais nous avons

$$\lim_{n \to \infty} \left[\sup_{\underset{\sim}{\theta} \in K_c} E_{P_n}(\varphi_n \mid \underset{\sim}{\theta}) - \operatorname*{Inf}_{\underset{\sim}{\theta} \in K_c} E_{P_n}(\varphi_n \mid \underset{\sim}{\theta}) \right] = 0$$

uniformément en c pour $\{\varphi_n\}$.

Donc de (1.4.4.) et (1.4.5.), il résulte qu'il existe une suite $\{c_{n'}\}$ telle que, pour $\underset{\sim}{\theta}_{n'} \in K_{c_{n'}}$.

$$E_{P_{n'}}(\varphi^*_{n'} \mid \underset{\sim}{\theta}_{n'}) > E_{P_{n'}}(\varphi_{n'} \mid \underset{\sim}{\theta}_{n'}) + \frac{\delta}{2}$$

pour n assez grand. Mais cela contredit la proposition 1.2.2. C.Q.F.D.

1.5. Régions de confiance

Dans ce paragraphe nous allons indiquer comment on peut se servir du test ci-dessus pour construire des régions de confiance optimum. Ce qui suit est un analogue multiparamétrique de la définition des régions de confiance les plus petites, donnée par Wald [53] .

Soit $\underset{\sim}{X}^{(n)} = (X_1, \ldots, X_n)$ une réalisation du processus. Par $S_n(\underset{\sim}{X}^{(n)})$, on note pour chaque n et chaque $\underset{\sim}{X}^{(n)}$ un sous-ensemble de Θ .

DÉFINITION 1.5.1. — Une suite de régions de confiance $\{S_n(\underset{\sim}{X}^{(n)})\}$ est dite asymptotiquement la plus petite, non biaisée pour $\underset{\sim}{\theta}$, si les conditions suivantes sont satisfaites

(1) $$P_n \left\{ \underset{\sim}{X}^{(n)} \;\middle|\; S_n(\underset{\sim}{X}^{(n)}) \ni \underset{\sim}{\theta} ; \underset{\sim}{\theta} \right\} = \alpha \quad , \quad \forall\, \underset{\sim}{\theta} \in \Theta$$

(2) $$\limsup_{n \to \infty} \left[\sup_{\underset{\sim}{\theta}', \underset{\sim}{\theta}''} \left\{ P_n \{ \underset{\sim}{X}^{(n)} \;\middle|\; S_n(\underset{\sim}{X}^{(n)}) \ni \underset{\sim}{\theta}' ; \underset{\sim}{\theta}'' \} \right\} \right] \leqslant \alpha$$

(condition d'absence de biais asymptotique).

(3) Pour chaque suite $\{S_n^*(\underset{\sim}{X}^{(n)})\}$ quelconque qui satisfait aux conditions (1) et (2) ,

$$\lim_{n \to \infty} \left\{ \sup_{\underset{\sim}{\theta}', \underset{\sim}{\theta}''} \left[P_n \{ X^{(n)} \;\middle|\; S_n(\underset{\sim}{X}^{(n)}) \ni \underset{\sim}{\theta}' ; \underset{\sim}{\theta}'' \} - P_n \{ \underset{\sim}{X}^{(n)} \;\middle|\; S_n^*(\underset{\sim}{X}^{(n)}) \ni \underset{\sim}{\theta}' ; \underset{\sim}{\theta}'' \} \right] \right\} = 0$$

α est appelé le coefficient de confiance de la région $S_n(\underset{\sim}{X}^{(n)})$.

LEMME 1.5.1. – *Soit* $\{W_n(\underset{\sim}{\theta})\}_{n \in \mathbf{N}^*}$ *pour chaque* $\underset{\sim}{\theta}$ *une suite d'ensembles dans l'espace des échantillons satisfaisant aux conditions suivantes*

(1) $$P_n(W_n(\underset{\sim}{\theta}) ; \underset{\sim}{\theta}) = 1 - \alpha \quad , \quad \forall\, \underset{\sim}{\theta} \in \Theta$$

(2) $$\sup_{\underset{\sim}{\theta}', \underset{\sim}{\theta}'' \in \Theta} P_n(W_n(\underset{\sim}{\theta}') ; \underset{\sim}{\theta}'') \longrightarrow 1 - \alpha \quad , \textit{ lorsque } n \to \infty .$$

(3) *Pour chaque suite* $\{W_n^*(\underset{\sim}{\theta})\}_{n \in \mathbf{N}^*}$ *quelconque qui satisfait aux conditions* (1) *et* (2)

$$\sup_{\underset{\sim}{\theta}', \underset{\sim}{\theta}''} \left[P_n(W_n(\underset{\sim}{\theta}') ; \underset{\sim}{\theta}'') - P_n(W_n^*(\underset{\sim}{\theta}') ; \underset{\sim}{\theta}'') \right] \to 0 \textit{ lorsque } n \to \infty .$$

Notons par $S_n(\underset{\sim}{X}^{(n)})$ *l'ensemble des valeurs* $\underset{\sim}{\theta} \in \Theta$ *tel que* $\underset{\sim}{X}^{(n)} \in \complement\, W_n(\underset{\sim}{\theta})$. *Alors on a*

(a) $$P_n \{ \underset{\sim}{X}^{(n)} \;\middle|\; S_n(\underset{\sim}{X}^{(n)}) \ni \underset{\sim}{\theta} ; \underset{\sim}{\theta} \} = \alpha , \quad \forall\, \theta \in \Theta$$

(b) $$\sup_{\underset{\sim}{\theta}', \underset{\sim}{\theta}'' \in \Theta} \left[P_n \{ \underset{\sim}{X}^{(n)} \;\middle|\; S_n(\underset{\sim}{X}^{(n)}) \ni \underset{\sim}{\theta}' ; \underset{\sim}{\theta}'' \} \right] \to \alpha \text{ lorsque } n \to \infty .$$

(c) *Pour chaque suite* $\{S_n^*(X^{(n)})\}_{n \in \mathbf{N}^*}$ *quelconque qui satisfait aussi aux conditions* (a) *et* (b)

$$\sup_{\underset{\sim}{\theta}', \underset{\sim}{\theta}'' \in \Theta} \left[P_n \{ \underset{\sim}{X}^{(n)} \mid S_n(\underset{\sim}{X}^{(n)}) \ni \underset{\sim}{\theta}' ; \underset{\sim}{\theta}'' \} - P_n \{ \underset{\sim}{X}^{(n)} \mid S_n^*(\underset{\sim}{X}^{(n)}) \ni \underset{\sim}{\theta}' ; \underset{\sim}{\theta}'' \} \right] \to 0 \quad \textit{lorsque } n \to \infty .$$

Ce lemme se démontre directement par l'identité

$$P_n \{ \underset{\sim}{X}^{(n)} \mid S_n(\underset{\sim}{X}^{(n)}) \ni \underset{\sim}{\theta}' ; \underset{\sim}{\theta}'' \} = 1 - P_n(W_n(\underset{\sim}{\theta}') ; \underset{\sim}{\theta}'') .$$

Maintenant, la région $\{W_n(\underset{\sim}{\theta})\}$ définie dans la proposition 1.2.2. satisfait aux conditions (1)-(3) du lemme 1.5.1. ci-dessus. Donc, la suite des régions $\{S_n(\underset{\sim}{X}^{(n)})\}_{n \in \mathbf{N}^*}$ ainsi construite possède les propriétés (a)-(c) de ce lemme. Par conséquent, $\{S_n(\underset{\sim}{X}^{(n)})\}_{n \in \mathbf{N}^*}$ est la suite de régions de confiance $\underset{\sim}{\theta}$ asymptotiquement la plus petite et sans biais.

Les régions sont écrites explicitement

$$S_n(\underset{\sim}{X}^{(n)}) = \{ \underset{\sim}{\theta} : Q_n \leqslant d_n \}$$

où Q_n est la forme quadratique et d_n le nombre choisi dans la proposition 1.2.2.

Remarque. — On peut remarquer que Wilks et Daly [57] ont utilisé la même statistique dans le cas de variables aléatoires indépendantes pour obtenir des régions de confiance asymptotiquement les plus petites. Pourtant, la définition "d'une suite de régions asymptotiquement les plus petites" est différente de celle qui est donnée ici. Leurs études ont comme objet de construire des régions de confiance optimales à partir des fonctions appartenant à une classe définie. D'après ce qui est fait ici, il s'en suit que les régions de confiance obtenues ici sont asymptotiquement optimales sans aucune condition sur la classe de fonctions, Cependant, comme la statistique appartient à la classe des fonctions considérées par Wilks et Daly, il en résulte que les régions de confiance sont "asymptotiquement optimales" même selon la définition donnée par ces derniers.

Le cas des échantillons finis a été traité par Bartlett [11] . Il a donné quelques corrections pour l'asymétrie et l'aplatissement. Il a démontré que par une application de ces corrections, on peut obtenir "des régions de confiance presque optimales" (optimun approximate confidence regions). Les mêmes raisonnements s'appliquent dans le cas présent.

1.6. Exemples

Dans ce paragraphe, le critère du test est calculé explicitement dans le cas (1) de la loi multinomiale et (2) de la loi gamma à trois paramètres.

1.6.1. *Critère du test pour la loi multinomiale*

Supposons que la population se répartisse en k classes différentes $C_1,\ldots,C_k$ dans les proportions $p_1,\ldots,p_k$ où $\sum_1^k p_i = 1$, $0 \leqslant p_i \leqslant 1$.

Notons $\underset{\sim}{p} = (p_1,\ldots,p_k)$. Les nombres p_i sont inconnus et on désire tester l'hypothèse

$$H_o : \underset{\sim}{p} = \underset{\sim}{p}_o = (p_{1o}, \ldots, p_{ko}) .$$

Un échantillon de taille n , $(X_1,\ldots,X_n)$ est tiré de la population et soient $\nu_1,\ldots,\nu_k$, avec $\sum_{i=1}^k \nu_i = n$ les nombres observées dans les k classes. C'est-à-dire $\nu_i = \sum_{j=1}^n I_{C_i} \circ X_j$ où I_{C_i} désigne la fonction indicatrice de C_i .

On sait que

$$\Pr\{(\nu_1,\ldots,\nu_k) = (n_1,\ldots,n_k) ; \underset{\sim}{p}_o\} = \frac{n\ !}{n_1\ !\ldots n_k\ !}\, p_{1o}^{n_1}\ldots p_{ko}^{n_k} = L_n$$

(disons)

Sans restreindre la généralité, on peut supposer que $p_i > 0$, $p_{io} > 0$, $\forall i = 1,\ldots,k$.

Puisque $\Sigma p_i = 1$, on a : $p_k = 1 - p_1 - \ldots - p_{k-1}$. Cela veut dire qu'il y a effectivement (k-1) paramètres inconnus.

La matrice d'information pour une seule observation est donnée par

$$\mathcal{J}(\underset{\sim}{p}) = \begin{bmatrix} \frac{1}{p_1} + \frac{1}{p_k}, \frac{1}{p_k}, \ldots, \frac{1}{p_k} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \frac{1}{p_k}, \ldots\ldots\ldots, \frac{1}{p_{k-1}} + \frac{1}{p_k} \end{bmatrix} .$$

$\mathcal{J}(\underset{\sim}{p})$ est une matrice d'ordre $(k-1) \times (k-1)$, symétrique et définie positive. Donc

$$\mathcal{J}^{-1}(\underset{\sim}{p}) = \begin{bmatrix} p_1(1-p_1), -p_1 p_2, \ldots, -p_1 p_{k-1} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ -p_1 p_{k-1}, \ldots\ldots\ldots, p_{k-1}(1-p_{k-1}) \end{bmatrix} = [I^{ij}(\underset{\sim}{p})]$$

Donc le critère du test pour l'hypothèse H_o basé sur un échantillon de taille n est

$$T_n(\underset{\sim}{X}^{(n)}) = \frac{1}{n} \sum_{i,j=1}^{k-1} I^{ij}(\underset{\sim}{p}_o) \left(\frac{\partial \log L_n}{\partial p_i}\right)_{\underset{\sim}{p}=\underset{\sim}{p}_o} \left(\frac{\partial \log L_n}{\partial p_j}\right)_{\underset{\sim}{p}=\underset{\sim}{p}_o}$$

Des calculs simples nous conduisent à

$$T_n(\underset{\sim}{X}^{(n)}) = \frac{1}{n} \sum_{i,j=1}^{k} J_{ij}(\underset{\sim}{p}_o) \frac{n_i}{p_{io}} \cdot \frac{n_j}{p_{jo}}$$

où $J(\underset{\sim}{p}) = [J_{ij}(\underset{\sim}{p})] = \begin{bmatrix} p_1(1-p_1), \ldots\ldots, -p_1 p_k \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ -p_1 p_k, \ldots\ldots, p_k(1-p_k) \end{bmatrix}$

est une matrice symétrique d'ordre $(k \times k)$. Donc

$$T_n = \sum_{i=1}^{k} \frac{p_{io}(1-p_{io})}{n} \cdot \frac{n_i^2}{p_{io}^2} - \sum_{\substack{i \neq j \\ i,j=1}}^{k} \frac{p_{io}\, p_{jo}}{n} \cdot \frac{n_i\, n_j}{p_{io} p_{jo}}$$

$$= \frac{1}{n} \left[\sum_{i=1}^{k} \frac{n_i^2}{p_{io}} - \sum_{i=1}^{k} n_i^2 - \sum_{\substack{i \neq j \\ i,j=1}}^{k} n_i\, n_j \right]$$

$$= \frac{1}{n} \left[\sum_{i=1}^{k} \frac{n_i^2}{p_{io}} - n^2 \right] .$$

Ainsi, dans ce cas particulier de la loi multinomiale, le critère T_n coïncide avec le critère de test du χ^2 (Test d'adéquation).

1.6.2. *Le critère du test dans le cas de la loi gamma à 3 paramètres.*

La loi gamma à trois paramètres est donnée [51] par la densité

$$f(x\,;\, a,b,\nu) = \frac{|b|}{a^{b\nu}\, \Gamma(\nu)} \left\{ \exp\left(\frac{-x}{a}\right) b \right\} x^{b\nu - 1}$$

pour $x \geqslant 0$, $b \neq 0$ et a et ν sont positifs.

Nous ne considérons que le cas où b aussi est positif. La loi gamma généralisée est un modèle fréquent dans la théorie de la fiabilité. Il est aussi démontré que la loi gamma généralisée engendre plusieurs lois de probabilités connues et couramment utilisées dans les applications, pour des valeurs différentes des paramètres a, b et ν . Par exemple, si $\nu = 1$, la loi est la loi de Weibull et pour $a = 2$, $\nu = \frac{n}{2}$ et $b = 1$, on retrouve la loi de χ^2 avec n degrés de liberté. La loi avec ses propriétés est étudiée en détail dans [37] et [51] . Il est remarqué aussi qu'il est extrêmement difficile d'obtenir les estimateurs du maximum de vraisemblance comme racines des équations de vraisemblance bien qu'elles existent. Donc, le critère que nous avons considéré donne une méthode pour les tests d'hypothèses concernant des paramètres sans calculer les estimateurs de paramètres.

La matrice d'information pour une seule observation est

$$\mathcal{J}(\underset{\sim}{\theta}) = \begin{bmatrix} \frac{1}{b^2}\left[1 + 2\psi(\nu) + \nu\psi^2(\nu) + \nu\psi'(\nu)\right] & \frac{-1}{a}\left[1 + \nu\psi(\nu)\right] & \frac{-\psi(\nu)}{b} \\ \frac{-1}{a}\left[1 + \nu\psi(\nu)\right] & \frac{b^2\nu}{a^2} & b/a \\ \frac{-\psi(\nu)}{b} & \frac{b}{a} & \psi'(\nu) \end{bmatrix}$$

où $\psi(\nu) = \frac{d}{d\nu} \operatorname{Log} \Gamma(\nu)$ et $\psi'(\nu) = \frac{d}{d\nu}\psi(\nu)$

Notons $\underset{\sim}{\theta} = (a, b, \nu)$

Donc

$$\mathcal{J}^{-1}(\underset{\sim}{\theta}) = \frac{1}{\Delta}\begin{bmatrix} \frac{b^2}{a^2}\left[\nu\psi'(\nu)-1\right] & \frac{1}{a}\left[\nu\psi(\nu)\psi'(\nu) + \psi'(\nu) - \psi(\nu)\right] & \frac{-b}{a^2} \\ \frac{1}{a}\left[\nu\psi(\nu)\psi'(\nu) + \psi'(\nu) - \psi(\nu)\right] & \frac{1}{b^2}\left[\psi'(\nu)(\nu\psi(\nu)-1) + \nu(\psi'(\nu))^2 + 2\psi(\nu)\psi'(\nu) + \psi(\nu)\right] & \frac{-1}{ba}\left[1 + \psi(\nu) + \nu\psi'(\nu)\right] \\ \frac{-b}{a^2} & \frac{-1}{ba}\left[1 + \psi(\nu) + \nu\psi'(\nu)\right] & \frac{1}{a^2}\left[-1 + \nu(\psi'(\nu) + 1)\right] \end{bmatrix}$$

où $\Delta = \text{dét } \mathcal{J}(\underset{\sim}{\theta}) = \frac{1}{a^2}\left[\nu(\psi(\nu))^2 - \psi'(\nu) - 1\right]$

Notons

$$X = \sum_{i=1}^{n} \log \left(\frac{X_i}{a} \right)$$

$$Y = \sum_{i=1}^{n} \frac{X_i}{a}^{b} \log \left(\frac{X_i}{a} \right)$$

$$Z = \sum_{i=1}^{n} \left(\frac{X_i}{a} \right)^{b} \quad ,$$

et L_n la fonction de vraisemblance pour un échantillon de taille n de la population. Alors

$$\frac{\partial \log L_n}{\partial a} = \frac{b}{a} Z - \frac{b \nu n}{a}$$

$$\frac{\partial \log L_n}{\partial b} = \frac{n}{b} + \nu X - Y$$

$$\frac{\partial \log L_n}{\partial \nu} = bX - \psi(\nu) \ .$$

Donc le critère du test est donné comme suit

$$T_n(\underset{\sim}{X}^{(n)}) = \frac{1}{n\Delta a^2} \Big\{ (n + b\nu X - bY) \left[bX \left(\nu^2 \psi(\nu) \right.\right.$$

$$- \nu - 1) + bY(1 - \nu\psi'(\nu)) + Z(\psi'(\nu) - \psi(\nu) - \nu\psi(\nu)\psi'(\nu))$$

$$+ \psi(\nu)(\nu^2 n \psi(\nu) + \nu n + 1) - n \Big]$$

$$+ (Z - \nu n) \left[bX \{ \nu^2 \psi(\nu)\psi'(\nu) - 2\nu\psi'(\nu) + \psi(\nu)(\nu - 1) - 1 \} \right.$$

$$+ bY \{ \psi'(\nu) - \psi(\nu) - \nu\psi(\nu)\psi'(\nu) \}$$

$$+ Z \{ \psi^2(\nu)(\nu\psi(\nu) - 1) + \nu(\psi'(\nu))^2 + 2\psi(\nu)\psi'(\nu) + \psi'(\nu) \}$$

$$+ n\nu \{ \psi'(\nu)(\nu\psi(\nu) - 1) - \nu(\psi'(\nu))^2 - \psi(\nu)\psi'(\nu) - \psi'(\nu) \}$$

$$+ (n+1) \{ \psi(\nu) + \psi'(\nu) + n\psi(\nu)\psi'(\nu) - n\psi'(\nu) \} \Big]$$

$$+ (bX - \psi(\nu)) \left[bX(\nu^2 \psi'(\nu) - 1) + bY \right.$$

$$+ Z(1 + \psi(\nu) + \nu\psi'(\nu)) + \nu^2 \psi'(\nu)(n - \psi'(\nu))$$

$$\left. + \psi(\nu)(n\nu + \nu + 1) + n(\nu - 1) \right] \Big\} \ .$$

Cependant, lorsque l'on remplace les paramètres par leurs valeurs correspondant à l'hypothèse nulle dans l'expression ci-dessus, celle-ci devient plus facile à manipuler. Par exemple, si l'on veut tester $H_o : \nu = 1, b = 1$ et $a = a_o$, ce qui revient à tester si la loi est exponentielle avec le paramètre a_o, la statistique T_n est un polynôme d'ordre deux en X, Y et Z donné par

$$
\begin{aligned}
T_n = \frac{1}{n\Delta a_o^2} \Big[& (n + X\text{-}Y) \{ X(\lambda - 2) + Y(1 - \lambda) \\
& + Z(\lambda - C + \lambda C) + C(n\lambda + n + 1) - n \} \\
& + (Z\text{-}n) \{ X(-\lambda C - 2\lambda - 1) + Y(\lambda + C - \lambda C) \\
& + Z[C^2(\lambda - 1) + \lambda^2 - 2C\lambda + \lambda] \\
& + n[C^2(\lambda - 1) + \lambda^2 + \lambda C - \lambda] + (n + 1)C + \lambda^2 - n\lambda C - n\lambda \} \\
& + (X + C) \{ X(\lambda - 1) + Y + Z(1 - C + \lambda) + \lambda(n\text{-}\lambda) - C(n + 2) \} \Big]
\end{aligned}
$$

$$
\begin{aligned}
= \frac{1}{n\Delta a_o^2} \Big[& X^2(2\lambda - 3) + Y^2(\lambda - 1) + Z^2 \{ \lambda^2 - C^2 + \lambda(C - 1) \} \\
& + 2XY(\lambda - 2) + X \{ n\lambda(1\text{-}C) - C(2n + 3) - n - \lambda^2 \} \\
& + Y \{ C(n\lambda + n + 1) + n \} \\
& + Z \{ nC^2(\lambda - 1) - n\lambda(\lambda + 2) + (n + 1)C + \lambda^2 \} \\
& + \lambda^2(n^2 - n - C) - C^2(n(\lambda - 1) + n + 1) \\
& - n(n-1)\lambda C + 2n^2\lambda - n^2 \Big]
\end{aligned}
$$

où $\lambda = \psi'(1)$ est la constante d'Euler et $C = -\psi(1)$.

2. Tests d'hypothèses composées

2.0. Introduction

Dans ce chapitre, nous allons étudier le problème des tests d'hypothèses concernant plusieurs paramètres lorsque l'hypothèse est composée. Il est bien connu que pour de tels problèmes, les régions semblables les plus puissantes n'existent que dans quelques cas particuliers. Dans le cas général, pour tester une hypothèse composée qui spécifie un paramètre les autres restant non-spécifiés, Neyman [39] a construit une classe appelée "classe $C(\alpha)$ des tests asymptotiques" basée sur des fonctions qui appartiennent à une classe appelée "classe des fonctions normalisées de Cramer" qui satisfait à certaines conditions de régularité et une classe des estimateurs pour des paramètres non-spécifiés satisfaisant à certaines propriétés assez faibles. On sait que la dérivée partielle du logarithme de la fonction de vraisemblance par rapport au paramètre soumis au test (à la valeur correspondant à H_o) fournit un test localement le plus puissant contre les alternatives unilatérales, et un test localement le plus puissant sans biais contre les alternatives bilatérales. Cependant, ce test comporte des paramètres inconnus. Cette fonction appartient à la "classe des fonctions de Cramer" et il est démontré qu'un test basé sur cette fonction est asymptotiquement optimal dans la classe $C(\alpha)$ quand on remplace des paramètres non-spécifiés par leur estimateur du type $\sqrt{n}$ convergent (voir Paragraphe 2.2. pour la définition). Dans le paragraphe 2.1., nous allons indiquer comment le résultat est valable pour le cas du processus de Markov à temps discret. Le paragraphe 2.2. donne quelques résultats nécessaires pour traiter le cas général de plusieurs paramètres.

Le paragraphe 2.3. est divisé en deux parties. La première partie traite du problème de l'hypothèse linéaire, tandis que la seconde traite du problème de l'hypothèse composée générale. Nous allons proposer un test pour de tels problèmes et nous étudions aussi ses propriétés. Il est démontré qu'avec certains types d'estimateurs substitués aux paramètres non spécifiés le test a asymptotiquement la meilleure puissance constante sur une famille de surfaces et est asymptotiquement le plus strict. On voit aussi que ce critère du test conduit, pour n assez grand, aux critères des tests proposés par Rao [45], Wald [54] et au critère du rapport de vraisemblance (pour de tels problèmes) en remplaçant des paramètres non-spécifiés par leurs différents estimateurs.

2.1. Cas d'un seul paramètre spécifié,

Sous les hypothèses du paragraphe 1.1., soit $p(\xi, \eta ; \gamma, \theta)$ la densité de la mesure de transition stationnaire qui contient $(k + 1)$ paramètres γ et $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)$. Supposons que les hypothèses 1.1. - 1.5 du chapitre 1 sont vérifiées dans ce cas avec $(k+1)$ paramètres. Supposons qu'on veut tester

$$H_o : \gamma = \gamma_o \ , \ \underset{\sim}{\theta} \text{ n'étant pas spécifié.}$$

Soit $\varphi(\xi, \eta ; \gamma , \underset{\sim}{\theta}) = \log p(\xi, \eta ; \gamma , \underset{\sim}{\theta})$. Notons par φ_γ et φ_{θ_i} les dérivées partielles de φ par rapport à γ et θ_i , $i = 1,\ldots,k$ .Notons également

$$\lambda_{\gamma\gamma}(\gamma, \underset{\sim}{\theta}) = E_{\gamma,\underset{\sim}{\theta}} \{ (\varphi_\gamma(X_1, X_2 ; \gamma, \underset{\sim}{\theta}))^2 \}$$

$$\lambda_{\gamma\theta_i}(\gamma, \underset{\sim}{\theta}) = E_{\gamma, \underset{\sim}{\theta}} \{ \varphi_\gamma(X_1, X_2 ; \gamma, \underset{\sim}{\theta}). \varphi_{\theta_i}(X_1, X_2 ; \gamma, \underset{\sim}{\theta}) \}$$

$$\lambda_{\theta_i\theta_j}(\gamma, \underset{\sim}{\theta}) = E_{\gamma, \underset{\sim}{\theta}} \{ \varphi_{\theta_i}(X_1, X_2 ; \gamma, \underset{\sim}{\theta}) \varphi_{\theta_j}(X_1, X_2 ; \gamma, \underset{\sim}{\theta}) \}$$

pour $i, j = 1,\ldots,k$, et

$$\Lambda = \begin{bmatrix} \lambda_{\gamma\gamma} & \lambda_{\gamma\theta_1} & \cdots & \lambda_{\gamma\theta_k} \\ \lambda_{\theta_1\gamma} & \lambda_{\theta_1\theta_1} & \cdots & \lambda_{\theta_1\theta_k} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_{\theta_k\gamma} & \cdots & \cdots & \lambda_{\theta_k\theta_k} \end{bmatrix}$$

décomposée en blocs de la façon suivante

$$\begin{bmatrix} \lambda_{\gamma\gamma} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}$$

où Λ_{12} est le vecteur-ligne $(\lambda_{\gamma\theta_1}, \ldots, \lambda_{\gamma\theta_k})$, Λ_{21} son transposé et Λ_{22} la matrice $[\lambda_{\theta_i\theta_j}]$ d'ordre $(k \times k)$. Enfin, notons par

(2.1.1.) $$(a_1(\gamma, \underset{\sim}{\theta}), \ldots, a_k(\gamma, \underset{\sim}{\theta})) = \Lambda_{12} \Lambda_{22}^{-1}$$

Soit

(2.1.2.) $$Y(\xi, \eta; \gamma_o, \underset{\sim}{\theta}) = \varphi_{\gamma_o} - \Sigma a_i(\gamma_o, \underset{\sim}{\theta}) \varphi_{\theta_i}$$

et soit $V(Y) = \sigma^2(\gamma_o, \underset{\sim}{\theta})$ la variance de Y .

Considérons la variable normalisée

(2.1.3.) $$Z(\xi, \eta; \gamma_o, \underset{\sim}{\theta}) = \frac{1}{\sqrt{V(Y)}} Y(\xi, \eta; \gamma_o, \underset{\sim}{\theta}) .$$

[Nous rappelons que les espérances mathématiques sont prises toujours avec la même notation que dans le chapitre 1] .

LEMME 2.1.1. – *Supposons qu'on ait un échantillon* $\underset{\sim}{X}^{(n)} = (X_1, \ldots, X_n)$ *du processus et soit*

(2.1.4.) $$Z_n(\underset{\sim}{X}^{(n)}; \gamma_o, \underset{\sim}{\theta}) = \frac{1}{\sqrt{n-1}} \sum_{i=1}^{n-1} Z(X_i, X_{i+1}; \gamma_o, \underset{\sim}{\theta})$$

Alors $Z_n(\underset{\sim}{X}^{(n)}; \gamma_o, \underset{\sim}{\theta})$ *suit asymptotiquement la loi normale d'espérance mathématique* 0 *et de variance unité.*

Ce résultat est obtenu par une application du théorême central limite de Lévy-Lindeberg pour les martingales comme dans la proposition 1.1.1.

Sous les hypothèses il s'en suit que les estimateurs du maximum de vraisemblance existent pour des paramètres non spécifiés avec une probabilité qui tend vers 1 quand $n \to \infty$; ils sont convergents presque sûrement et suivent asymptotiquement une loi normale (cf. ch.5). En outre les $\lambda_{..}$ sont continus en γ, $\underset{\sim}{\theta} \in \Theta$. Supposons que les paramètres inconnus sont remplacés par leurs estimateurs du maximum de vraisemblance, $\hat{\underset{\sim}{\theta}}_n$, racines des équations de vraisemblance.

LEMME 2.1.2. –

(1) $$Z_n(\underset{\sim}{X}^{(n)}; \gamma_o, \hat{\underset{\sim}{\theta}}_n) = \frac{1}{\sqrt{(n-1)} \cdot \sigma(\gamma_o, \hat{\underset{\sim}{\theta}}_n)} \sum_{i=1}^{n-1} \varphi_{\gamma_o}(X_i, X_{i+1}; \gamma_o, \hat{\underset{\sim}{\theta}}_n)$$

suit asymptotiquement une loi normale d'espérance mathématique 0 *et de variance unité.*

(2) $Z_n(\underset{\sim}{X}^{(n)}; \gamma_o, \hat{\underset{\sim}{\theta}}_n)$ *fournit un test asymptotiquement optimal pour* H_o *dans la classe* $C(\alpha)$ *des tests.*

Le lemme se démontre de façon analogue au cas des variables aléatoires indépendantes traité par Neyman [ibid] et nous ne le faisons pas ici.

2.2. Quelques résultats généraux

Soit $p(\xi, \eta ; \underset{\sim}{\theta})$ la densité de transition avec le paramètre $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k, \theta_{k+1}, \ldots, \theta_{k+r})$ élément d'un ensemble compact Θ de $\mathbf{R}^{(k+r)}$.

Notons par $\underset{\sim}{\theta} = ({}_1\underset{\sim}{\theta}, {}_2\underset{\sim}{\theta})$ une partition générique de $\underset{\sim}{\theta}$ selon les nécessités. Nous notons également $\varphi(\xi, \eta ; \underset{\sim}{\theta}) = \log p(\xi, \eta ; \underset{\sim}{\theta})$ et φ_{θ_i} les dérivées partielles de $\varphi(\xi, \eta ; \underset{\sim}{\theta})$ par rapport à θ_i , i = 1,...,k+r . Soit

(2.2.1.) $\lambda_{ij}(\underset{\sim}{\theta}) = E_{\underset{\sim}{\theta}}(\varphi_{\theta_i}(X_1, X_2 ; \underset{\sim}{\theta})\ \varphi_{\theta_j}(X_1, X_2 ; \underset{\sim}{\theta}))$

et $\Lambda = [\lambda_{ij}(\underset{\sim}{\theta})]$, une matrice d'ordre $(k+r) \times (k+r)$.

Nous supposons comme ci-dessus, que $\Lambda(\underset{\sim}{\theta})$ est définie positive et que les hypothèses 1.1. - 1.6. du chapitre 1 sont remplies.

Considérons le problème du test d'hypothèse

(2.2.2.) $H_\omega : {}_1\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k) = (\theta_1^\circ, \ldots, \theta_k^\circ) = {}_1\underset{\sim}{\theta}^\circ$

${}_2\underset{\sim}{\theta} = (\theta_{k+1}, \ldots, \theta_{k+r})$ non spécifiés.

Notons $\underset{\sim}{\theta} = ({}_1\underset{\sim}{\theta}, {}_2\underset{\sim}{\theta})$ avec la partition ci-dessus.

Considérons

(2.2.3.) $Y_i(\xi, \eta ; \underset{\sim}{\theta}) = \varphi_{\theta_i}(\xi, \eta ; \underset{\sim}{\theta}) - \sum_{j=k+1}^{k+r} a_{ij}(\underset{\sim}{\theta})\, \varphi_{\theta_j}(\xi, \eta ; \underset{\sim}{\theta})$

pour $i = 1,\ldots,k$ où les $a_{ij}(\underset{\sim}{\theta})$ sont les coefficients définis ci-dessous. Supposons que

$$(2.2.4.) \qquad \Lambda(\underset{\sim}{\theta}) = \begin{pmatrix} \Lambda_{11}(\underset{\sim}{\theta}) & \Lambda_{12}(\underset{\sim}{\theta}) \\ \Lambda_{21}(\underset{\sim}{\theta}) & \Lambda_{22}(\underset{\sim}{\theta}) \end{pmatrix} ,$$

la matrice décomposée en blocs telle que Λ_{11} est la sous-matrice des k premières lignes et k premières colonnes de Λ et Λ_{22} est la sous-matrice correspondant aux éléments des r dernières lignes et r dernières colonnes de Λ . Puisque Λ est définie positive, Λ_{11} et Λ_{22} le sont aussi. Avec cette notation, la matrice des coefficients

$$(2.2.5.) \qquad (a_{ij}(\underset{\sim}{\theta}))_{(k\times r)} = \Lambda_{12}\,\Lambda_{22}^{-1} .$$

Soit $\quad \underset{\sim}{Y}^T(\xi, \eta\,;\,\underset{\sim}{\theta}) = ({}_1\underset{\sim}{Y}^T(\xi, \eta\,;\,\underset{\sim}{\theta}),\ {}_2\underset{\sim}{Y}^T(\xi, \eta\,;\,\underset{\sim}{\theta}))$

où $\quad {}_1\underset{\sim}{Y}^T(\xi, \eta\,;\,\underset{\sim}{\theta}) = (Y_1(\xi, \eta\,;\,\underset{\sim}{\theta}),\ldots,\ Y_k(\xi, \eta\,;\,\underset{\sim}{\theta}))$

$${}_2\underset{\sim}{Y}^T(\xi, \eta\,;\,\underset{\sim}{\theta}) = (\varphi_{\theta_{k+1}}(\xi, \eta\,;\,\underset{\sim}{\theta}),\ldots,\ \varphi_{\theta_{k+r}}(\xi, \eta\,;\,\underset{\sim}{\theta})) .$$

Alors ${}_1\underset{\sim}{Y}$ et ${}_2\underset{\sim}{Y}$ sont orthogonaux et leur matrice de covariance est donnée par

$$(2.2.6.) \qquad \Sigma(\theta) = \begin{pmatrix} \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\,\Lambda_{21} & 0 \\ 0 & \Lambda_{22} \end{pmatrix}$$

Les éléments en seront notés V_{ij} , $i,j = 1,\ldots,k+r$.

LEMME 2.2.1. — *Supposons qu'on ait un échantillon* $\underset{\sim}{X}^{(n)} = (X_1,\ldots,X_n)$ *de taille* n *du processus.*

Soit

$$(2.2.7.) \qquad Z_i^{(n)}(\underset{\sim}{Z}^{(n)}\,;\,\underset{\sim}{\theta}) = \frac{1}{\sqrt{n-1}} \sum_{j=1}^{n-1} Y_i(X_j, X_{j+1}\,;\,\underset{\sim}{\theta})$$

pour $i = 1,\ldots,k$.

Alors $\underset{\sim}{Z}^{(n)} = (Z_1^{(n)}, \ldots, Z_k^{(n)})^T$ *suit asymptotiquement la loi normale de vecteur espérance mathématique* $\underset{\sim}{\theta}$ *et de matrice de covariances*

$$A = \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21} .$$

La démonstration de ce lemme se fait comme celle du lemme 2.1.1., en appliquant le théorème central limite de Lévy-Lindeberg pour les martingales.

D'où le corollaire suivant

COROLLAIRE. – La forme quadratique

$$(2.2.8.) \qquad Q_n = \underset{\sim}{Z}^{(n)T} \; A^{-1} \; \underset{\sim}{Z}^{(n)}$$

suit asymptotiquement une loi de χ^2 *avec* k *degrés de liberté.*

(Il est clair que A est définie positive et donc A^{-1} existe. Cela résulte aussi du lemme 2.2.3.) .

Comme dans le lemme 2.1.2., supposons que les paramètres $_2\underset{\sim}{\theta}$ sont remplacés par leurs estimateurs " $\sqrt{n}$-convergents " $_2\underset{\sim}{\hat{\theta}}_n$. Alors, on a le résultat suivant

LEMME 2.2.2. – $\underset{\sim}{Z}^{(n)}(\underset{\sim}{X}^{(n)}; {}_1\underset{\sim}{\theta}^{\circ}, {}_2\underset{\sim}{\hat{\theta}}_n)$ *suit asymptotiquement la même loi que* $\underset{\sim}{Z}^{(n)}(\underset{\sim}{Z}^{(n)}; {}_1\underset{\sim}{\theta}^{\circ}, {}_2\underset{\sim}{\theta})$, *c'est-à-dire la loi normale avec vecteur espérance mathématique* 0 *et matrice de covariance* A , *puisque les* $\lambda_{ij}(\underset{\sim}{\theta})$ *sont uniformément continus en* $\underset{\sim}{\theta} \in \Theta$. *Donc, on a aussi que la forme quadratique*

$$(2.2.9.) \quad Q(\underset{\sim}{X}^{(n)}; {}_1\underset{\sim}{\theta}^{\circ}, {}_2\underset{\sim}{\hat{\theta}}_n) = \underset{\sim}{Z}^{(n)T}(\underset{\sim}{X}^{(n)}; {}_1\underset{\sim}{\theta}^{\circ}, {}_2\underset{\sim}{\hat{\theta}}_n) A^{-1} ({}_1\underset{\sim}{\theta}^{\circ}, {}_2\underset{\sim}{\hat{\theta}}_n) \cdot$$

$$\cdot \; \underset{\sim}{Z}^{(n)}(\underset{\sim}{X}^{(n)}; {}_1\underset{\sim}{\theta}^{\circ}, {}_2\underset{\sim}{\hat{\theta}}_n)$$

suit asymptotiquement une loi de χ^2 *avec* k *degrés de liberté.*

Comme le signalent Neyman et Scott [41] , lorsque l'on considère les hypothèses composées, une des principales difficultés est de trouver

une classe de tests qui conserve, au moins asymptotiquement, le seuil donné. Ceci nous conduit à exiger que, sous l'hypothèse soumise au test, le critère du test suive asymptotiquement une loi qui ne dépende pas des paramètres nuisibles.

On peut noter que la forme quadratique Q donnée par (2.2.8.) satisfait à ces exigences lorsqu'on remplace les paramètres non-spécifiés par leurs estimateurs "$\sqrt{n}$-convergents". Donc on peut l'utiliser pour tester les hypothèses composées.

Rappelons la définition d'un estimateur "$\sqrt{n}$-convergent" donnée par Neyman [39]

DÉFINITION. – On dit que les estimateurs $\hat{\theta}_{1n},\ldots,\hat{\theta}_{rn}$ des paramètres $\theta_1,\ldots,\theta_r$ constituent un système faiblement $\sqrt{n}$-convergent, si pour tout ensemble $a_1,\ldots,a_r$ de nombres réels tels que $\sum_1^r a_i^2 > 0$, $\sqrt{n}\sum_{i=1}^r a_i(\hat{\theta}_{in} - \theta_i)$ ne tend pas en probabilité vers 0 .

En particulier si $(\sqrt{n}(\hat{\theta}_{1n} - \theta_1),\ldots,\sqrt{n}(\hat{\theta}_{rn} - \theta_r)$ suit asymptotiquement une loi non dégénérée, alors les estimateurs $(\hat{\theta}_{1n},\ldots,\hat{\theta}_{rn})$ constituent un système faiblement $\sqrt{n}$-convergent .

Maintenant, les estimateurs du maximum de vraisemblance racines convergentes des équations correspondantes, forment un système faiblement $\sqrt{n}$-convergent dans le cas considéré. Donc, si l'on remplace $_2\underset{\sim}{\theta}$ par $_2\underset{\sim}{\hat{\theta}}_n$ estimateurs du maximum de vraisemblance, racines des équations correspondant à la valeur $_1\underset{\sim}{\theta}^\circ$ pour les autres, la forme quadratique devient

$$(2.2.10)\qquad Q(\underset{\sim}{X}^{(n)};\ {}_1\underset{\sim}{\theta}^\circ,\ {}_2\underset{\sim}{\hat{\theta}}_n) = \frac{1}{(n-1)}\left(\frac{\partial \log L_n}{\partial \theta_1},\ldots,\frac{\partial \log L_n}{\partial \theta_k}\right) A^{-1} \begin{pmatrix} \frac{\partial \log L_n}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log L_n}{\partial \theta_k} \end{pmatrix}$$

où L_n est la densité de l'échantillon et cette expression est calculée en $({}_1\underset{\sim}{\theta}^\circ,\ {}_2\underset{\sim}{\hat{\theta}}_n)$ et $A^{-1} = A^{-1}({}_1\underset{\sim}{\theta}^\circ,\ {}_2\underset{\sim}{\hat{\theta}}_n)$.

Remarquons qu'on a retrouvé le critère du test proposé par Rao [44] pour $H\omega$. Plus tard, on verra qu'en remplaçant $\underset{\sim}{\theta}_2$ par d'autres estimateurs, le test possède quelques propriétés optimales comme dans le cas d'hypothèses simples.

Le lemme qui suit est une conséquence du théorème de Jacobi sur les mineurs de la matrice adjointe ([4] p. 98).

Soit $\mathcal{B} = [b_{ij}]$ une matrice d'ordre $(s \times s)$ symétrique et définie positive. Notons

$$\mathcal{B}^{-1} = [e_{ij}] .$$

Soit

$$\mathcal{B} = \begin{bmatrix} \mathcal{B}_{11} & \mathcal{B}_{12} \\ \mathcal{B}_{21} & \mathcal{B}_{22} \end{bmatrix}$$

une partition de $\mathcal{B}$ telle que

$$\mathcal{B}_{11} = \begin{pmatrix} b_{11},\ldots, b_{1t} \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots \\ b_{t1},\ldots, b_{tt} \end{pmatrix} \quad \text{et} \quad \mathcal{B}_{22} = \begin{pmatrix} b_{t+1\ t+1},\ldots, b_{t+1\ s} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ b_{s\ t+1},\ldots, b_{ss} \end{pmatrix}$$

Notons par $\mathcal{B}_{22}^{-1} = (d_{ij})$, $i,j = t+1,\ldots,s$.

LEMME 2.2.3. –

$$\mathcal{B}_{11} - \mathcal{B}_{12}\,\mathcal{B}_{22}^{-1}\,\mathcal{B}_{21} = (e_{ij})^{-1}_{i,j=1,\ldots,t} = (c_{ij}) .$$

DÉMONSTRATION. – Soit $A = \mathcal{B}_{11} - \mathcal{B}_{12}\,\mathcal{B}_{22}^{-1}\,\mathcal{B}_{21} = (a_{m\,n})$.

Alors, $a_{mn} = b_{mn} - \sum_{i=t+1}^{s} \sum_{j=t+1}^{s} d_{ij}\, b_{im}\, b_{jn}$

Pour plus de facilité, considérons a_{11}

$$a_{11} = b_{11} - \sum_{i=t+1}^{s} \sum_{j=t+1}^{s} d_{ij}\ b_{i1}\ b_{1j} =$$

$$= \begin{vmatrix} b_{11}, b_{1\,t+1}, b_{1\,t+2}, \ldots, b_{1s} \\ b_{t+1\,1}, b_{t+1\,t+1}, \ldots, b_{t+1\,s} \\ b_{s1}, b_{s\,t+1}, \ldots, b_{ss} \end{vmatrix} \Bigg\backslash \begin{vmatrix} b_{t+1\,t+1}, \ldots, b_{t+1\,s} \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots \\ b_{s\,t+1}, \ldots, b_{ss} \end{vmatrix}$$

Le théorème de Jacobi affirme que si H est une matrice d'ordre s (nous supposons qu'elle est symétrique et définie positive ; mais ceci n'est pas nécessaire pour l'énoncé du théorème de Jacobi), D est un mineur d'ordre t de H et D* est le mineur d'ordre t correspondant dans (Adj H). Alors

$$D^* = (\text{dét } H)^{t-1} \quad (\text{cofacteur de D dans H}) \ .$$

Pour appliquer ce théorème, soit

$$H = (e_{ij})_{s \times s} \ ; \ (\text{Adj H}) = (\text{dét H})\, \mathcal{B} \ .$$

Soit

$$D = \begin{bmatrix} e_{11}, e_{1\,t+1}, \ldots, e_{1s} \\ e_{t+1\,1}, e_{t+1\,t+1}, \ldots, e_{t+1\,s} \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots \\ e_{s1}, e_{s\,t+1}, \ldots, e_{ss} \end{bmatrix}$$

Alors $D^* = (\text{dét H})^{s-t+1} \begin{vmatrix} b_{11}, b_{1\,t+1}, \ldots, b_{1s} \\ b_{t+1\,1}, b_{t+1\,t+1}, \ldots, b_{t+1\,s} \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots \\ b_{s1}, b_{s\,t+1}, \ldots, b_{ss} \end{vmatrix}$

$$= (\text{dét } H)^{s-t+1-1} \begin{vmatrix} e_{22} , \dots , e_{2t} \\ \dots\dots\dots\dots \\ \dots\dots\dots\dots \\ e_{t2} , \dots , e_{tt} \end{vmatrix}$$

De même si $E = \begin{vmatrix} e_{t+1\ t+1} , \dots , e_{t+1\ s} \\ \dots\dots\dots\dots\dots\dots \\ \dots\dots\dots\dots\dots\dots \\ e_{st+1} , \dots , e_{ss} \end{vmatrix}$

$$E^* = (\text{dét } H)^{s-t} \begin{vmatrix} b_{t+1\ t+1} , \dots , b_{t+1\ s} \\ \dots\dots\dots\dots\dots\dots \\ b_{st+1} , \dots , b_{ss} \end{vmatrix}$$

$$= (\text{dét } H)^{s-t-1} \begin{vmatrix} e_{11} & \dots & e_{1t} \\ \dots\dots\dots\dots\dots \\ e_{t1} & \dots & e_{tt} \end{vmatrix}$$

Par conséquent,

$$a_{11} = \begin{vmatrix} e_{22} & \dots & e_{2t} \\ e_{t2} & \dots & e_{tt} \end{vmatrix} \Big/ \begin{vmatrix} e_{11} & \dots & e_{1t} \\ e_{t1} & \dots & e_{tt} \end{vmatrix} = c_{11}$$

De même, $a_{mn} = c_{mn}$, $\forall\, m, n = 1, \dots, t$

où $c_{ij} = (e_{ij})^{-1}$ $i, j = 1, \dots, t$.

En appliquant ce lemme avec $\mathcal{B} = \Lambda$, si l'on note
$\Lambda^{-1} = (\lambda^{ij})_{i,j\,=\,1,\dots,\,k+r}$, la forme quadratique (2.2.8.) peut s'écrire

(2.2.11.) $Q = \underset{\sim}{Z}^{(n)T} [(\lambda^{ij})_{i,j\,=\,1\dots k}]^{-1} \underset{\sim}{Z}^{(n)}$

où $(\lambda^{ij})_{i,j=1,\dots,k}$ est la sous-matrice de Λ^{-1} correspondant à $i, j = 1, \dots, k$.

2.3. Tests d'hypothèses composées concernant plusieurs paramètres.

2.3.A. *Tests d'hypothèses composées linéaires.*

Soit ω un sous-ensemble de Θ , ensemble de paramètres et l'on considère l'hypothèse nulle $H\omega : \underset{\sim}{\theta} \in \omega$. Un point de ω sera noté par $\overline{\underset{\sim}{\theta}}$ pour le distinguer du point $\underset{\sim}{\theta} \in \Theta - \omega$.

DÉFINITION 2.3.1. (Wald). — Une région critique $W_n \in \mathcal{S}^{(n)}$ pour tester l'hypothèse $H\omega$ est dite avoir uniformément la meilleure puissance constante sur une famille de surfaces $K_c(\overline{\underset{\sim}{\theta}})$ si les conditions suivantes sont remplies

(1) $$P_n(W_n ; \underset{\sim}{\theta}') = P_n(W_n ; \underset{\sim}{\theta}'')$$

pour tout couple $\underset{\sim}{\theta}'$, $\underset{\sim}{\theta}''$ de points appartenant à la même surface $K_c(\overline{\underset{\sim}{\theta}})$;

(2) $$P_n(W_n ; \underset{\sim}{\theta}) \geqslant P_n(W_n^* ; \underset{\sim}{\theta})$$

pour chaque $\underset{\sim}{\theta}$ qui n'appartient pas à ω et pour chaque région critique $W_n^* \in \mathcal{S}^{(n)}$ telle que

$$\sup_{\overline{\underset{\sim}{\theta}}} P_n(W_n ; \overline{\underset{\sim}{\theta}}) = \sup_{\overline{\underset{\sim}{\theta}}} P_n(W_n^* ; \overline{\underset{\sim}{\theta}}) .$$

La deuxième partie de la condition (2) exprime que la région W_n^* a le même seuil que W_n .

Considérons un vecteur aléatoire $\underset{\sim}{U}^T = (U_1,\ldots,U_k)$ qui suit une loi normale à k dimensions de vecteur espérance mathématique $\underset{\sim}{\theta}^T = (\theta_1,\ldots,\theta_k)$ inconnu et de matrice de covariance $\Lambda^{-1} = (\lambda_{ij})^{-1}$ définie positive connue. Soit $\varphi_{\theta_i}(\underset{\sim}{U} ; \underset{\sim}{\theta})$, la dérivée partielle du logarithme de la densité par rapport à θ_i, $i = 1,\ldots,k$. Les deux lemmes suivants se vérifient directement.

LEMME 2.3.1. — Le *vecteur* $(\varphi_{\theta_1},\ldots,\varphi_{\theta_k})$ *suit aussi la loi normale d'espérance mathématique* 0 *et de matrice de covariance* Λ .

LEMME 2.3.1. – *Soient*

$$Z_i(\underset{\sim}{U}\ ;\ \underset{\sim}{\theta}) = \varphi_{\theta_i}(\underset{\sim}{U}\ ;\ \underset{\sim}{\theta}) - \sum_{j=1}^{k-r} a_{ij+r}(\underset{\sim}{\theta})\,\varphi_{\theta_{j+r}}(\underset{\sim}{U}\ ;\ \underset{\sim}{\theta})$$

pour $i = 1,\ldots,r$ $(r < k)$ *où les* $a_{ij}(\underset{\sim}{\theta})$ *sont des coefficients de régression partielle de* φ_{θ_i} *en* φ_{θ_j} .

Alors $\underset{\sim}{Z} = (Z_1,\ldots, Z_r)$ *suit une loi normale d'espérance mathématique* 0 *et de matrice de covariances*

$$A = [(\lambda^{ij})_{i,j=1,\ldots,r}]^{-1} \quad où \quad (\lambda^{ij}) = \Lambda^{-1}$$

Considérons le problème : tester l'hypothèse
$H\omega : \theta_1 = \theta_1^\circ\ ,\ldots,\ \theta_r = \theta_r^\circ$ $(r < k)$ et les $\theta_{r+1},\ldots,\theta_k$ ne sont pas spécifiés, c'est-à-dire ω est un sous-ensemble de Θ dont les points ont les r premières coordonnées $= (\theta_1^\circ,\ldots,\theta_r^\circ) = ({}_1\underset{\sim}{\theta}^\circ)$. Notons $\overline{\underset{\sim}{\theta}} \in \omega$ par $({}_1\underset{\sim}{\theta}^\circ\ ,\ {}_2\overline{\underset{\sim}{\theta}})$.

Considérons un échantillon $\underset{\sim}{U}^{(n)} = (\underset{\sim}{U}_1,\ldots,\underset{\sim}{U}_n)$ de taille n extrait de la population ; soit

$$(2.3.1.)\quad Z_i^{(n)} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_i(\underset{\sim}{U}_j\ ;\ {}_1\underset{\sim}{\theta}^\circ\ ;\ {}_2\underset{\sim}{\theta})$$

pour $i = 1,\ldots,r$ et soit $\underset{\sim}{Z}^{(n)} = (Z_1^{(n)},\ldots, Z_r^{(n)})^T$.

Alors, il est clair que $\underset{\sim}{Z}^{(n)}$ suit la loi normale d'espérance mathématique 0 et de matrice de covariance $A = [(\lambda^{ij})_{i,j=1,\ldots,r}]^{-1}$. Donc il résulte que

$$Q(\underset{\sim}{Z}^{(n)}\ ;\ \overline{\underset{\sim}{\theta}}) = \underset{\sim}{Z}^{(n)T} A^{-1}\, \underset{\sim}{Z}^{(n)}$$

suit la loi de χ^2 à r degrés de liberté.

Maintenant, supposons que les paramètres non spécifiés sont remplacés par leurs estimateurs qui constituent un système faiblement $\sqrt{n}$-convergent. Alors la loi de $Q(\underset{\sim}{U}^{(n)}\ ;\ {}_1\underset{\sim}{\theta}^\circ,\ {}_2\overline{\underset{\sim}{\theta}})$ est asymptotiquement celle d'un χ^2 avec r degrés de liberté.

Donc, nous avons une famille de tests pour l'hypothèse $H\omega$. Supposons en particulier que ${}_2\underset{\sim}{\theta}$ sont remplacés par leurs estimateurs du maximum de vraisemblance (sans restriction), racines des équations de vraisemblance. Alors, on voit que la forme quadratique suit une loi de χ^2 à r degrés de liberté. On démontre la proposition suivante

PROPOSITION 2.3.1. – *La région critique*

$$(2.3.2.) \qquad W : \left\{ \underset{\sim}{U} \;\middle|\; \underset{\sim}{Z}^T A^{-1} \underset{\sim}{Z} \geqslant d \right\}$$

basée sur un seul échantillon avec ${}_2\underset{\sim}{\theta}$ *remplacé comme ci-dessus et* d *choisi pour satisfaire à la condition* $\sup_{\overline{\underset{\sim}{\theta}}} P(W ; \overline{\underset{\sim}{\theta}}) = \alpha$ *condition de seuil, a uniformément la meilleure puissance constante sur une famille de surfaces* $K_c(\overline{\underset{\sim}{\theta}})$ *définie par*

$$(2.3.3.) \qquad K_c(\overline{\underset{\sim}{\theta}}) : ({}_1\underset{\sim}{\theta} - {}_1\underset{\sim}{\theta}^\circ)^T A ({}_1\underset{\sim}{\theta} - {}_1\underset{\sim}{\theta}^\circ) = c$$

$$\gamma_{t1}\theta_1 + \ldots + \gamma_{tk}\theta_k = \sum_{i=1}^{k} \gamma_{ti}\overline{\theta}_i$$

où les γ_{ti} *sont définis au cours de la démonstration (pour éviter la répétition) et où* $\overline{\underset{\sim}{\theta}} \in \omega$ *est dans ce qui suit, un point fixé arbitrairement.*

DÉMONSTRATION. –

$$(2.3.4.) \qquad \begin{aligned} W &= \{ \underset{\sim}{U} \mid \underset{\sim}{Z}^T A^{-1} \underset{\sim}{Z} \geqslant d \} \\ &= \{ \underset{\sim}{U} \mid ({}_1\underset{\sim}{U} - {}_1\underset{\sim}{\theta}^\circ) A ({}_1\underset{\sim}{U} - {}_1\underset{\sim}{\theta}^\circ) \geqslant d \} \end{aligned}$$

comme on le voit après quelques calculs en écrivant les expressions explicitement et en remplaçant les composantes de ${}_2\underset{\sim}{\theta}$ par leur estimateur du maximum de vraisemblance (sans restriction).

Nota. – Cette réduction nous donne déjà une démonstration intuitive pour le lemme (2.2.3.), puisque $({}_1\underset{\sim}{U} - {}_1\underset{\sim}{\theta}^\circ)$ suit la loi normale de vecteur espérance mathématique 0 et de matrice de covariance A^{-1} (sous $H\omega$) .

c'est-à-dire

$$W : \{ \underset{\sim}{U} \mid ({}_1\underset{\sim}{U} - {}_1\underset{\sim}{\theta}^\circ)^T (\lambda^{ij})^{-1} ({}_1\underset{\sim}{U} - {}_1\underset{\sim}{\theta}^\circ) \geqslant d \} .$$

Donc la condition (1) est satisfaite pour $K_c(\underset{\sim}{\overline{\theta}})$.

Maintenant, considérons la transformation linéaire non singulière dans l'espace des échantillons comme suit

$$y_i - \theta_i^\circ = \beta_{i1}(U_1 - \theta_1^\circ) + \ldots + \beta_{ir}(U_r - \theta_r^\circ)$$

pour $i = 1,\ldots,r$,

(2.3.5.) $y_i = \gamma_{i1}U_1 + \ldots + \gamma_{ik}U_k$ pour $i = r+1,\ldots, k$, telle que les variables $(y_1,\ldots,y_k) = \underset{\sim}{Y} = ({}_1\underset{\sim}{Y}, {}_2\underset{\sim}{Y})$ soient indépendantes et suivent une loi normale de variance unité. La région critique est donnée par

(2.3.6.) $S : -\{ \underset{\sim}{Y} \mid ({}_1\underset{\sim}{Y} - {}_1\underset{\sim}{\theta}^\circ)^T \ ({}_1\underset{\sim}{Y} - {}_1\underset{\sim}{\theta}^\circ) \geqslant d \}$.

Considérons la même transformation pour les paramètres

(2.3.7.) $\theta_i' - \theta_i^\circ = \beta_{i1}(\theta_1 - \theta_1^\circ) + \ldots + \beta_{ir}(\theta_r - \theta_r^\circ)$

pour $i = 1,\ldots,r$.

$\theta_i' = \gamma_{i1}\theta_1 + \ldots + \gamma_{ik}\theta_k \quad i = r+1,\ldots, k$.

Alors, la surface $K_c(\underset{\sim}{\overline{\theta}})$ se transforme en

(2.3.8.) $K_c'(\underset{\sim}{\overline{\theta}}) = \sum_{i=1}^{r} (\theta_i' - \theta_i^\circ)^2 = c$

$\theta_i' = \sum_{j=1}^{k} \gamma_{ij}\overline{\theta}_j$ pour $i = r+1,\ldots, k$,

c'est-à-dire qu'avec cette transformation, pour tout point $\underset{\sim}{\theta}' \in K_c'(\underset{\sim}{\theta})$, ${}_2\underset{\sim}{\theta}'$ est constant.

Comme dans le cas correspondant de l'hypothèse simple, supposons qu'on note les probabilités après les transformations, par $P'(\underset{\sim}{Y} ; \underset{\sim}{\theta}')$ avec la densité

(2.3.9.) $p'(\underset{\sim}{Y} ; \underset{\sim}{\theta}') = p'_1({}_1\underset{\sim}{Y} ; {}_1\underset{\sim}{\theta}')\, p'_2({}_2\underset{\sim}{Y} ; {}_2\underset{\sim}{\theta}')$,

puisque les variables sont indépendantes.

Si l'on démontre que la région S est la plus puissante pour H contre une alternative convenablement choisie, alors le résultat est démontré comme dans la proposition 1.2.1.

Considérons la mesure alternative avec la densité

$$\int_{K'_c} p'(\underset{\sim}{Y} ; \underset{\sim}{\theta}')\,dA = \int_{K'_c} p'_1({}_1\underset{\sim}{Y} ; {}_1\underset{\sim}{\theta}')\, p'_2({}_2\underset{\sim}{Y} ; {}_2\underset{\sim}{\theta}')\,dA =$$

$$= p'_2({}_2\underset{\sim}{Y}' ; {}_2\underset{\sim}{\theta}') \int_{K'_c} p'_1({}_1\underset{\sim}{Y} ; {}_1\underset{\sim}{\theta}')\,dA$$

car ${}_2\underset{\sim}{\theta}'$ est constant sur K'_c .

On a également $p'(\underset{\sim}{Y} ; \overline{\underset{\sim}{\theta}}') = p'_1({}_1\underset{\sim}{Y} ; {}_1\underset{\sim}{\theta}^\circ)\, p_2({}_2\underset{\sim}{Y} ; {}_2\underset{\sim}{\theta}')$ avec la même ${}_2\underset{\sim}{\theta}'$ comme ci-dessus. Pour démontrer que S est la plus puissante pour H'_ω contre cette alternative, il suffit de démontrer que

$$(2.3.2.)\qquad \frac{\int_{K'_c} p'(\underset{\sim}{Y} ; \underset{\sim}{\theta}')dA}{p'(\underset{\sim}{Y} ; \overline{\underset{\sim}{\theta}}')} \begin{cases} \geqslant d & \text{dans S} \\ < d & \text{dehors} \end{cases}$$

c'est-à-dire

$$\frac{p'_2({}_2\underset{\sim}{Y} ; {}_2\underset{\sim}{\theta}') \int_{K'_c} p'_1({}_1\underset{\sim}{Y} ; {}_1\underset{\sim}{\theta}')\,dA}{p'_2({}_2\underset{\sim}{Y} ; {}_2\underset{\sim}{\theta}')\, p'_1({}_1\underset{\sim}{Y} ; {}_1\underset{\sim}{\theta}^\circ)} \begin{cases} \geqslant d & \text{dans S} \\ < d & \text{dehors.} \end{cases}$$

Soit

$$\frac{\int_{K'_c} p'_1({}_1\underset{\sim}{Y} ; {}_1\underset{\sim}{\theta}')\,dA}{p'_1({}_1\underset{\sim}{Y} ; {}_1\underset{\sim}{\theta}^\circ)} \begin{cases} \geqslant d & \text{dans S} \\ < d & \text{dehors.} \end{cases}$$

Mais cela est le cas de l'hypothèse simple et (2.3.2.) se vérifie donc comme dans la proposition 1.2.1. Par conséquent, la proposition est démontrée.

Au cours de la démonstration, il est démontré que le test est optimal d'après la définition 2.3.1., même dans une classe plus élargie pour inclure certains tests randomisés.

DÉFINITION 2.3.2. (Wald). — On dit qu'une suite $\{W_n\}_{n \in \mathbf{N}^*}$ de régions critiques pour tester l'hypothèse H ω a asymptotiquement la meilleure puissance constante sur une famille de surfaces $K_c(\underset{\sim}{\theta})$ si les conditions suivantes sont remplies

(1) $$\sup_{\underset{\sim}{\theta} \in \omega} P_n(W_n ; \underset{\sim}{\theta}) = \alpha \ , \ \forall n \in \mathbf{N}^*$$

(2) $$\lim_{n \to \infty} \left[\sup_{c, \underset{\sim}{\theta}} \left\{ \sup_{\underset{\sim}{\theta} \in K_c(\underset{\sim}{\theta})} P_n(W_n ; \underset{\sim}{\theta}) - \inf_{\underset{\sim}{\theta} \in K_c(\underset{\sim}{\theta})} P_n(W_n ; \underset{\sim}{\theta}) \right\} \right] = 0$$

(3) Pour chaque suite $\{W_n^*\}_{n \in \mathbf{N}^*}$ de régions critiques pour lesquelles les conditions (1) et (2) sont réalisées ,

$$\lim_{n \to \infty} \left\{ \sup_{\underset{\sim}{\theta} \in \complement \omega} \left[P_n(W_n ; \underset{\sim}{\theta}) - P_n(W_n^* ; \underset{\sim}{\theta}) \right] \right\} = 0$$

où $\complement \omega$ est le complément de ω dans Θ .

Maintenant, nous nous plaçons dans le cas du paragraphe 2.2. et de l'hypothèse nulle y donnée. Nous avons vu que la forme quadratique Q donnée par (2.2.8.) peut s'employer comme un critère du test avec des paramètres non spécifiés, remplacés par leurs estimateurs qui constituent un système faiblement $\sqrt{n}$-convergent .

Considérons la région

(2.3.12) $$W_n : \{ \underset{\sim}{X}^{(n)} \in \mathcal{S}^{(n)} \mid Q(\underset{\sim}{X}^{(n)} ; {}_1\underset{\sim}{\theta}^{\circ}, {}_2\underset{\sim}{\hat{\theta}}_n) \geqslant d_n \}$$

où $Q(. ; .)$ est la forme quadratique définie par (2.2.9.) et les estimateurs ${}_2\underset{\sim}{\hat{\theta}}_n$ remplacés sont des estimateurs du maximum de vraisemblance (sans restriction) et d_n est choisi pour satisfaire à la condition (1) de la définition 2.3.2.

C'est-à-dire

$$\sup_{\underset{\sim}{\theta}} P_{\cdot}(W_n ; \underset{\sim}{\theta}) = \alpha \qquad \forall\, n \in \mathbf{N}^*$$

Considérons la transformation linéaire non singulière

$$(2.3.13) \qquad V_i(\xi, \eta ; \underset{\sim}{\theta}) = \beta_{i1}(\underset{\sim}{\theta}) Y_1(\xi, \eta ; \underset{\sim}{\theta}) + \ldots + \beta_{ik}(\underset{\sim}{\theta}) Y_k(\xi, \eta ; \underset{\sim}{\theta})$$

pour $i = 1,\ldots,k$

et

$$V_i(\xi, \eta ; \underset{\sim}{\theta}) = \gamma_{i1}(\underset{\sim}{\theta})\, Y_1(\xi, \eta ; \underset{\sim}{\theta}) + \ldots + \gamma_{ik+r}(\underset{\sim}{\theta})\, Y_{k+r}(\xi, \eta ; \underset{\sim}{\theta})$$

pour $i = k+1,\ldots,k+r$

telle que la matrice des covariances de $\underset{\sim}{V}$ est la matrice identité d'ordre (k+r), c'est-à-dire si

$$(2.3.14) \qquad \mathcal{B}(\underset{\sim}{\theta}) = \begin{bmatrix} \beta_{11}(\underset{\sim}{\theta}), \ \ldots\ , \beta_{1k}(\underset{\sim}{\theta}) , \ 0\ , \ldots\ , \ 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \beta_{k1}(\underset{\sim}{\theta}), \ \ldots\ , \beta_{kk}(\underset{\sim}{\theta}) , \ 0\ , \ldots\ , \ 0 \\ \gamma_{k+1,1}(\underset{\sim}{\theta}), \ldots, \ \ldots\ldots\ldots\ldots, \gamma_{k+1,k+r}(\underset{\sim}{\theta}) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \gamma_{k+r,1}(\underset{\sim}{\theta}), \ldots, \ \ldots\ldots\ldots\ldots, \gamma_{k+r,k+r}(\underset{\sim}{\theta}) \end{bmatrix}$$

et

$$(2.3.15) \qquad \Sigma(\underset{\sim}{\theta}) = \begin{bmatrix} A(\underset{\sim}{\theta}) & 0 \\ 0 & \Lambda_{22}(\underset{\sim}{\theta}) \end{bmatrix},$$

la matrice des covariances de $\underset{\sim}{Y}$; alors

(2.3.16) $\mathcal{B}^T(\underset{\sim}{\theta})\,\Sigma(\underset{\sim}{\theta})\,\mathcal{B}(\underset{\sim}{\theta}) = I$ la matrice identité.

Par conséquent, il s'en suit que $\mathcal{B}(\underset{\sim}{\theta})$ est non-singulière. Puisque $\Sigma(\underset{\sim}{\theta})$ est continue et bornée dans Θ, $\mathcal{B}(\underset{\sim}{\theta})$ l'est aussi. Considérons un point $\overline{\underset{\sim}{\theta}} \in \omega$ fixé, noté par $({}_1\underset{\sim}{\theta}^\circ, {}_2\overline{\underset{\sim}{\theta}})$ et la surface $K_{c_n}(\overline{\underset{\sim}{\theta}})$ définie par

$$K_{c_n}(\overline{\underset{\sim}{\theta}}): ({}_1\underset{\sim}{\theta} - {}_1\underset{\sim}{\theta}^\circ)^T A(\overline{\underset{\sim}{\theta}})({}_1\underset{\sim}{\theta} - {}_1\underset{\sim}{\theta}^\circ) = c_n \quad \text{où} \quad c_n \to o$$

quand $n \to \infty$.

(2.3.17) $$\sum_{i=1}^{k+r} \gamma_{ti}(\overline{\underset{\sim}{\theta}})\,\theta_i = \sum_{i=1}^{k+r} \gamma_{ti}(\overline{\underset{\sim}{\theta}})\,\overline{\theta}_i$$

pour $t = k+1, \ldots, k+r$.

Il en résulte qu'il existe un nombre positif δ tel que, pour chaque point $\underset{\sim}{\theta}$ tel que $|\theta_i - \theta_i^\circ| \leqslant \delta$ pour $i = 1,\ldots,k$, il existe exactement un point $\overline{\underset{\sim}{\theta}} \in \omega$ et une valeur c telle que $\underset{\sim}{\theta} \in K_c(\overline{\underset{\sim}{\theta}})$ (Wald [54]) .

Considérons la transformation des paramètres

(2.3.18) $$\theta'_i - \theta_i^\circ = \beta_{i1}(\overline{\underset{\sim}{\theta}})(\theta_1 - \theta_1^\circ) + \ldots + \beta_{ik}(\overline{\underset{\sim}{\theta}})(\theta_k - \theta_k^\circ)$$

pour $i = 1,\ldots,k$

$$\theta'_i = \gamma_{i1}(\overline{\underset{\sim}{\theta}})\,\theta_1 + \ldots + \gamma_{ik+r}(\overline{\underset{\sim}{\theta}})\,\theta_{k+r}$$

pour $i = k+1, \ldots, k+r$.

Alors, la surface $K_{c_n}(\overline{\underset{\sim}{\theta}})$ définie par (2.3.17) est transformée en

$$K'_{c_n}(\overline{\underset{\sim}{\theta}}): \sum_{1}^{k} (\theta'_i - \theta_i^\circ)^2 = c_n$$

$$\theta'_t = \sum_{i=1}^{k+r} \gamma_{ti}(\overline{\underset{\sim}{\theta}})\theta_i = \sum_{i=1}^{k+r} \gamma_{ti}(\overline{\underset{\sim}{\theta}})\,\overline{\theta}_i$$

$t = k+1,\ldots,k+r$

c'est-à-dire pour tous les points $\underset{\sim}{\theta}' \in K'_{c_n}(\overline{\underset{\sim}{\theta}})$ les coordonnées $(\theta'_{k+1}, \ldots, \theta'_{k+r})$ sont les mêmes.

PROPOSITION 2.3.2. — *La suite* $\{W_n\}_{n \in \mathbf{N}^*}$ *de régions critiques définies par* (2.3.12) *(pour tester l'hypothèse* $H\omega$*) pour chaque* n, *a asymptotiquement la meilleure puissance constante sur la famille des surfaces* $K_{c_n}(\underset{\sim}{\theta})$ *définie par* (2.3.17).

DÉMONSTRATION. — La condition (1) est vérifiée par la construction de W_n, c'est-à-dire par le choix de d_n pour chaque n. Puisque $\underset{\sim}{Z}^{(n)}(\underset{\sim}{\theta})$ suit asymptotiquement une loi normale uniformément en $\underset{\sim}{\theta}$ avec la matrice de covariances $A(\underset{\sim}{\theta})$, il résulte que

$$\lim_{n \to \infty} \left[\sup_{\underset{\sim}{\theta} \in K_{c_n}(\overline{\underset{\sim}{\theta}})} P_n(W_n ; \underset{\sim}{\theta}) - \inf_{\underset{\sim}{\theta} \in K_{c_n}(\overline{\underset{\sim}{\theta}})} P_n(W_n ; \underset{\sim}{\theta}) \right] = 0$$

et par conséquent la condition (2) est vérifiée par $\{W_n\}$. Pour démontrer que la condition (3) est satisfaite, en vertu de la remarque 1.2.1., on peut se placer dans l'espace image par $\underset{\sim}{Z}^{(n)}$. Puisque $\underset{\sim}{Z}^{(n)}$ suit asymptotiquement une loi normale, la proposition se démontre si, pour une suite $\{W_n^*\}_{n \in \mathbf{N}^*}$ quelconque (nous utilisons le même symbole pour une région dans l'espace des échantillons que dans l'espace image quand il n'y a pas d'ambiguité) qui satisfait aux conditions (1) et (2),

on a

$$\lim_{n \to \infty} \left\{ \sup_{\theta \in \mathbb{C}\ \omega} [\Phi(W_n^* ; \underset{\sim}{\theta}) - \Phi(W_n ; \underset{\sim}{\theta})] \right\} = 0$$

où $\Phi(. ; \underset{\sim}{\theta})$ est la loi gaussienne de vecteur espérance mathématique 0 et de matrice de covariances $A(\underset{\sim}{\theta})$. Pour $\underset{\sim}{\theta}$ tel que $|\theta_i - \theta_i^\circ| > \delta$ pour $i = 1,\ldots,k$ où $\delta > 0$ est le nombre obtenu ci-dessus, on a

$$\lim_{n \to \infty} \Phi(W_n ; \underset{\sim}{\theta}) = 1$$

comme dans la proposition 1.3.2.

On peut donc considérer l'espace des paramètres où

$$|\theta_i - \theta_i^\circ| < \delta \qquad \forall\ i = 1,\ldots,k \ .$$

Si la proposition n'est pas vraie, il existe une suite $\{\underset{\sim}{\theta}^n\}$ telle que

$$\lim_{n \to \infty} \left\{ \Phi (W_n^* ; \underset{\sim}{\theta}^n) - \Phi (W_n ; \underset{\sim}{\theta}^n) \right\} = \varepsilon > 0 .$$

Pour chaque n, supposons que $\underset{\sim}{\theta}^n \in K_{c_n} (\overline{\underset{\sim}{\theta}})$, déterminé uniquement. Si $nc_n \to \infty$ alors $\Phi (W_n ; \underset{\sim}{\theta}^n) \to 1$ et par conséquent, il suffit de considérer les points tels que nc_n reste borné lorsque $n \to \infty$. Ainsi la condition de compacité de Θ n'est pas restrictive. Pour chaque n, considérons la régions $R_n (\overline{\underset{\sim}{\theta}})$ définie par

$$(2.3.19.) \quad \left\{ \underset{\sim}{X}^{(n)} \;\middle|\; \underset{\sim}{Z}^{(n)T}(\underset{\sim}{X}^{(n)}; {}_1\underset{\sim}{\theta}^\circ, {}_2\hat{\underset{\sim}{\theta}}_n) A^{-1}(\overline{\underset{\sim}{\theta}}) \underset{\sim}{Z}^{(n)}(\underset{\sim}{X}^{(n)}; {}_1\underset{\sim}{\theta}^\circ, {}_2\overline{\underset{\sim}{\theta}}_n) \geqslant e_n \right\}$$

où $\overline{\underset{\sim}{\theta}}$ est le point qui détermine $K_{c_n} (\overline{\underset{\sim}{\theta}})$ contenant le point $\underset{\sim}{\theta}^n$ et e_n est choisi pour satisfaire à la condition de seuil. Alors on a

$$(2.3.20) \quad \lim_{n \to \infty} [P_n(R_n (\overline{\underset{\sim}{\theta}}) ; \underset{\sim}{\theta}^n) - P_n(W_n ; \underset{\sim}{\theta}^n)] = 0 .$$

De même,

$$(2.3.21) \quad \lim_{n \to \infty} [P_n(R_n (\overline{\underset{\sim}{\theta}}) ; \underset{\sim}{\theta}^n) - \Phi(R_n (\overline{\underset{\sim}{\theta}}) ; \underset{\sim}{\theta}^n)] .$$

Donc, si la proposition n'est pas exacte, il en résulte que

$$(2.3.22) \quad \lim_{n \to \infty} [\Phi (W_n^* ; \underset{\sim}{\theta}^n) - \Phi (R_n (\overline{\underset{\sim}{\theta}}) ; \underset{\sim}{\theta}^n)] = \varepsilon > 0 ,$$

ce qui contredit la proposition 1.1. D'où le résultat.

Dans la statistique $Q(\underset{\sim}{X}^{(n)} ; {}_1\underset{\sim}{\theta}^\circ, {}_2\hat{\underset{\sim}{\theta}}) =$

$$= \underset{\sim}{Z}^{(n)T}(\underset{\sim}{X}^{(n)} ; {}_1\underset{\sim}{\theta}^\circ, {}_2\hat{\underset{\sim}{\theta}}_n) \; A^{-1}({}_1\underset{\sim}{\theta}^\circ, {}_2\hat{\underset{\sim}{\theta}}_n) \underset{\sim}{Z}^{(n)}(\underset{\sim}{X}^{(n)} ; {}_1\underset{\sim}{\theta}^\circ, {}_2\hat{\underset{\sim}{\theta}}_n) ,$$

au lieu de remplacer seulement ${}_2\underset{\sim}{\theta}$ par ses estimateurs du maximum de vraisemblance dans $A^{-1}({}_1\underset{\sim}{\theta}, {}_2\underset{\sim}{\theta})$, si l'on remplace $\underset{\sim}{\theta}$ par ses estimateurs du maximum de vraisemblance, on retrouve la statistique de Wald [54] pour de tels problèmes. D'ailleurs, cette statistique est asymptotiquement équivalente à la statistique du rapport de vraisemblance. Ainsi,

on peut considérer $Q(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta})$ comme une généralisation des statistiques courantes, pour un choix convenable des estimateurs. En même temps, on a un choix pour les paramètres non-spécifiés.

Tests les plus stricts pour les hypothèses linéaires composées.

Notons

$$M_n(\underset{\sim}{\theta}, \psi : \alpha) = \sup_{S_n} P_n(S_n ; \underset{\sim}{\theta})$$

où

$$S_n \in \mathcal{S}^{(n)} \quad \text{et} \quad \sup_{\overline{\underset{\sim}{\theta}} \in \omega} P_n(S_n ; \overline{\underset{\sim}{\theta}}) = \alpha .$$

DÉFINITION 2.3.3. (Wald). – Une région critique S_n est dite la plus stricte pour l'hypothèse $H\omega$ si

(1) $$\sup_{\overline{\underset{\sim}{\theta}}} P_n(S_n ; \overline{\underset{\sim}{\theta}}) = \alpha$$

(2) $$\sup_{\underset{\sim}{\theta}} \{ M_n(\underset{\sim}{\theta}, \omega : \alpha) - P_n(S_n ; \underset{\sim}{\theta}) \}$$

$$\leq \sup_{\underset{\sim}{\theta}} \{ M_n(\underset{\sim}{\theta}, \omega : \alpha) - P_n(S_n^* ; \underset{\sim}{\theta}) \}$$

pour chaque région S_n^* satisfaisant à la condition (1) .

PROPOSITION 2.3.3. – *La région critique définie dans la proposition* 2.3.1. *par* (2.3.2.) *est la plus stricte pour* $H\omega$ *considérée dans cette proposition*

DÉMONSTRATION. – La démonstration est due à Wald [54] puisque la statistique coïncide avec celle qui a été considérée par lui.

La démonstration consiste à montrer que $M(\underset{\sim}{\theta}, \omega : \alpha)$ est constant sur $K_c(\overline{\underset{\sim}{\theta}})$ définie par (2.3.3.) .

Considérons les transformations (2.3.13) et soit

$$M'(\underset{\sim}{\theta}', \omega' : \alpha) = \sup_{S'} P'(S' ; \underset{\sim}{\theta}')$$

où S' sont tels que $\sup_{\overline{\underset{\sim}{\theta}}' \in \omega'} P'(S' ; \overline{\underset{\sim}{\theta}}') = \alpha$

c'est-à-dire $\sup_{{}_2\underset{\sim}{\theta}' \in \omega'} P'(S' ; {}_1\underset{\sim}{\theta}^\circ, {}_2\underset{\sim}{\theta}') = \alpha$.

Alors $M(\underset{\sim}{\theta}, \omega : \alpha) = M'(\underset{\sim}{\theta}', \omega' : \alpha)$

où $\underset{\sim}{\theta}'$ est l'image de $\underset{\sim}{\theta}$.

Donc, il suffit de démontrer que $M'(\underset{\sim}{\theta}', \omega' : \alpha)$ est constant sur $K'_c(\overline{\underset{\sim}{\theta}})$ donné par (2.3.8.), c'est-à-dire

$$\sum_{i=1}^{r} (\theta'_i + \theta^\circ_i)^2 = c$$

$$\theta'_t = \sum_{i=1}^{k} \gamma_{ti} \overline{\theta}_i \qquad t = r+1,\ldots,k \ .$$

Soit $M^*(\underset{\sim}{\theta}', {}_1\underset{\sim}{\theta}^\circ : \alpha) = \sup_{W'} P'(W' ; \underset{\sim}{\theta}')$ pour lesquelles

$P'(W' ; {}_1\underset{\sim}{\theta}^\circ, {}_2\underset{\sim}{\theta}') = \alpha$.

Alors $M^*(\underset{\sim}{\theta}', {}_1\underset{\sim}{\theta}^\circ : \alpha) \geqslant M'(\underset{\sim}{\theta}', \omega' : \alpha)$.

Puisque les variables $Y_1,\ldots,Y_k$ sont indépendantes, on voit (par le lemme de Neyman-Pearson) que la région V' pour laquelle

(2.3.23) $\quad P'(V' ; {}_1\underset{\sim}{\theta}^\circ, {}_2\underset{\sim}{\theta}') = M^*(\underset{\sim}{\theta}', {}_1\underset{\sim}{\theta}^\circ : \alpha)$

est un sous-ensemble dans l'espace de $Y_1,\ldots,Y_r$ (de la forme $A^{(r)} \otimes \mathcal{Y}^{(k-r)}$ où $(\mathcal{Y}, \mathcal{C})$ est l'espace des Y et $A^{(r)} \in \mathcal{C}^{(r)}$) .

D'où $P'(V' ; {}_1\underset{\sim}{\theta}^\circ, {}_2\underset{\sim}{\theta}'^1) = P'(V' ; {}_1\underset{\sim}{\theta}^\circ, {}_2\underset{\sim}{\theta}'^2)$

pour tout couple $\underset{\sim}{\theta}'^1, \underset{\sim}{\theta}'^2 \in K'_c(\overline{\underset{\sim}{\theta}})$.

Donc $M^*(\underset{\sim}{\theta}', {}_1\underset{\sim}{\theta}^\circ : \alpha) = M'(\underset{\sim}{\theta}', \omega' : \alpha) = M(\theta, \omega : \alpha)$.

Maintenant, $M^*(\underset{\sim}{\theta}', {}_1\underset{\sim}{\theta}^\circ ; \alpha)$ étant constant sur $K'_c(\underset{\sim}{\theta})$, la démonstration est complétée par la proposition 2.3.1.

DÉFINITION 2.3.4. (Wald). — Une suite de régions critiques $\{S_n\}_{n \in \mathbf{N}^*}$ est dite asymptotiquement la plus stricte pour Hω si les deux conditions suivantes sont réalisées

(1) $$\sup_{\overline{\underset{\sim}{\theta}} \in \omega} P_n(S_n ; \overline{\underset{\sim}{\theta}}) = \alpha \qquad \forall\, n \in \mathbf{N}^*$$

(2) Pour chaque suite $\{S_n^*\}_{n \in \mathbf{N}^*}$ quelconque qui satisfait à la condition (1)

$$\limsup_{n \to \infty} \left\{ \sup_{\underset{\sim}{\theta} \in \Theta} [M_u(\underset{\sim}{\theta}, \omega : \alpha) - P_n(S_n ; \underset{\sim}{\theta})] - \sup_{\underset{\sim}{\theta} \in \Theta} [M_n(\underset{\sim}{\theta}, \omega : \alpha) - P_n(S_n^* ; \theta)] \right\} \leqslant 0$$

PROPOSITION 2.3.4. — *La suite des régions critiques définie par* (2.3.14) *est asymptotiquement la plus stricte pour* Hω .

La démonstration se fait comme pour la proposition précédente, $\underset{\sim}{Z}^{(n)}$ étant une variable aléatoire qui suit asymptotiquement la loi normale.

2. 3. B. *Hypothèses composées générales.*

Les fonctions vérifiables. — Linnik [35] a introduit le concept suivant de la vérifiabilité des fonctions de paramètres. Soit X une variable aléatoire dont la probabilité est donnée par $P(. ; \underset{\sim}{\theta})\ \underset{\sim}{\theta} \in \Theta$. Alors une fonction $\lambda(\underset{\sim}{\theta})$ de $\underset{\sim}{\theta}$ est dite vérifiable s'il existe un test $\varphi(X)$ pour $\lambda(\underset{\sim}{\theta})$ tel que $E_{\underset{\sim}{\theta}}(\varphi(X))$ n'est pas une constante et ne dépend de $\underset{\sim}{\theta}$ que par les valeurs de $\lambda(\underset{\sim}{\theta})$. Cependant, il n'existe pas de caractérisation générale de cette classe de fonctions vérifiables et il est démontré (Lehmann [34], Linnik [35]) qu'en général, la classe est assez restrictive. Donc, nous

ne considérons ici que l'approche classique consistant à réduire l'hypothèse composée générale convenablement comme une hypothèse linéaire par l'introduction de fonctions auxiliaires des paramètres, ce qui induit un homéomorphisme de l'espace des paramètres. Il est démontré que le test construit dans le paragraphe précédent peut aussi s'employer pour de tels problèmes. On démontre quelques propriétés optimales asymptotiques de ce test lorsqu'on utilise des estimateurs du maximum de vraisemblance pour $\underset{\sim}{\theta}$.

Supposons que l'hypothèse $H\omega$ est donnée par

(2.3.26) $\quad H\omega : \; U_1(\underset{\sim}{\theta}) = 0, \ldots, U_k(\underset{\sim}{\theta}) = 0$

Faisons l'hypothèse supplémentaire suivante

Hypothèse A. – Il existe r fonctions $U_{k+1}(\underset{\sim}{\theta}),\ldots,U_{k+r}(\underset{\sim}{\theta})$ satisfaisant aux conditions suivantes

(1) l'application $\underset{\sim}{\theta} : (\theta_1,\ldots,\theta_{k+r}) \longleftrightarrow (U_1,\ldots,U_{k+r}) = \underset{\sim}{U}$ est un homéomorphisme de l'espace des paramètres.

(2) La première et la seconde dérivées partielles de $\underset{\sim}{U}$ par rapport à θ_i , i = 1,...,k+r , existent et sont continues et bornées. Θ étant compact, elles sont uniformément continues en $\underset{\sim}{\theta}$.

(3) $$\inf_{\underset{\sim}{\theta} \in \Theta} \left| \frac{\partial(U_1, \ldots, U_{k+r})}{\partial(\theta_1, \ldots, \theta_{k+r})} \right| > 0$$

Notons par $P^*(\xi, . ; \underset{\sim}{U})$ et $P^*(. ; \underset{\sim}{U})$ respectivement la mesure de transition stationnaire et la mesure absolue stationnaire, avec les paramètres transformés. Ces mesures satisfont aussi aux hypothèses 1.1. - 1.6. du chapitre 1 chaque fois qu'il en est ainsi des mesures originales.

Soit

$$\lambda^*_{ij}(\underset{\sim}{\theta}) = - E_{\underset{\sim}{U}} \left(\frac{\partial^2}{\partial U_i \, \partial U_j} \log p^*(X_1, X_2 ; \underset{\sim}{U}) \right) =$$

$$= E_{\underset{\sim}{U}} \left(\frac{\partial}{\partial U_i} \log p^*(X_1, X_2 ; \underset{\sim}{U}) \; \frac{\partial}{\partial U_j} \log p(X_1, X_2 ; \underset{\sim}{U}) \right).$$

Alors

$$\lambda^*_{ij}(\theta) = \sum_{l=1}^{k+r} \sum_{m=1}^{k+r} \lambda_{lm}(\theta) \frac{d\theta_l}{dU_i} \frac{d\theta_m}{dU_j}$$

Soit $D = \left[\frac{\partial \theta_l}{\partial U_m}\right]_{l,m = 1,\ldots,k+r}$.

Alors $\Lambda^* = [\lambda^*_{ij}(\underset{\sim}{\theta})] = D^T \Lambda D$ où D^T est la transposée de D. D étant non singulière

$$\Lambda^{*-1} = [\lambda^{*ij}(\underset{\sim}{\theta})] = D^{-1} \Lambda^{-1} (D^T)^{-1}$$

c'est-à-dire $\lambda^{*ij}(\underset{\sim}{\theta}) = \sum_{l=1}^{k+r} \sum_{m=1}^{k+r} \lambda^{lm}(\underset{\sim}{\theta}) \frac{\partial U_i}{\partial \theta_l} \cdot \frac{\partial U_j}{\partial \theta_m}$

pour tout i,j = 1,...,k+r. .

Soit $\underset{\sim}{Z}^*(\xi, \eta ; \underset{\sim}{U})$, le vecteur analogue de $\underset{\sim}{Z}(\xi, \eta ; \theta)$ avec des paramètres transformés. Aussi, notons par $A^*(\underset{\sim}{\theta})$ sa matrice de covariances.

Puisque l'application $\underset{\sim}{\theta} \longleftrightarrow \underset{\sim}{U}$ est un homéomorphisme, on peut résoudre les équations

$$U_i = U_i(\underset{\sim}{\theta}) \qquad i = 1,\ldots,k+r \qquad \text{pour obtenir}$$

$$\theta_i = \theta_i(\underset{\sim}{U}) \qquad i = 1,\ldots,k+r \ .$$

Considérons la région critique donnée par

$$(2.3.27)\quad W_n : \left\{ \underset{\sim}{X}^{(n)} \;\middle|\; Q_n = \underset{\sim}{Z}^{*(n)\,T}(\underset{\sim}{X}^{(n)} ; {}_1\underset{\sim}{U}, {}_2\underset{\sim}{U}) A^{*-1}({}_1\underset{\sim}{U}, {}_2\underset{\sim}{U}) \cdot \underset{\sim}{Z}^{*(n)}(\underset{\sim}{X}^{(n)} ; {}_1\underset{\sim}{U}, {}_2\underset{\sim}{U}) \geqslant d_n \right\} .$$

Si $\hat{\underset{\sim}{U}}_n$ sont des estimateurs du maximum de vraisemblance de $\underset{\sim}{U}$, alors par l'hypothèse A, les estimateurs du maximum de vraisemblance de $\underset{\sim}{\theta}$ sont les $\theta(\hat{\underset{\sim}{U}}_n) = \hat{\underset{\sim}{\theta}}_n$.

Si l'on remplace ${}_2\underset{\sim}{U}$ par ses estimateurs du maximum de vraisemblance sous Hω, racines des équations, la forme quadratique devient

(2.3.28) $$Q_n(\underset{\sim}{X}^{(n)} ;\ {}_1\underset{\sim}{U}^{\circ},\ {}_2\underset{\sim}{\hat{U}}_n) =$$

$$= \frac{1}{n-1} \left(\frac{\partial \log L_n^*}{\partial U_1}, \ldots, \frac{\partial \log L_n^*}{\partial U_k} \right) A^{*-1} \begin{bmatrix} \frac{\partial \log L_n^*}{\partial U_1} \\ \vdots \\ \frac{\partial \log L_n^*}{\partial U_k} \end{bmatrix}$$

on peut la réduire comme suit

$$\varphi^*_{U_i} = \frac{\partial \log p^*(\xi, \eta ; \underset{\sim}{U})}{\partial U_i} = \sum_{j=1}^{k+r} \frac{\partial \log p\ (\xi, \eta\ ;\ \underset{\sim}{\theta})}{\partial \theta_j} \quad \frac{\partial \theta_j}{\partial U_i}$$

pour $i = 1,\ldots,k+r$.

D'où $$\begin{vmatrix} \varphi^*_{U_1} \\ \vdots \\ \varphi_{U_{k+r}} \end{vmatrix} = D \begin{vmatrix} \varphi_{\theta_1} \\ \vdots \\ \varphi_{\theta_{k+r}} \end{vmatrix}$$

ou $$\begin{vmatrix} \varphi^*_{U_1} \\ \vdots \\ \varphi^*_{U_k} \end{vmatrix} = {}_1D \begin{vmatrix} \varphi_{\theta_1} \\ \vdots \\ \varphi_{\theta_{k+r}} \end{vmatrix}$$

où ${}_1D$ est la sous-matrice d'ordre $k \times (k+r)$, sous-matrice des K premières lignes de D . Et puisque

$$\Lambda^{*-1} = D^{-1}\, \Lambda^{-1}\, (D^T)^{-1} \text{ , on a}$$

$$A^* = {}_1B\, \Lambda^{-1}\, {}_1B^T$$

où ${}_1B$ est la sous-matrice des k premières lignes de D^{-1} et ${}_1B^T$ sa transposée.

Donc, la forme quadratique Q peut s'écrire

$$(2.3.28')\quad \frac{1}{n-1}\left(\frac{\partial \log L_n}{\partial \theta_1},\ldots,\frac{\partial \log L_n}{\partial \theta_{k+r}}\right)_1 D^T\,({}_1B\,\Lambda^{-1}\,{}_1B^T)\,{}_1D \begin{vmatrix} \frac{\partial \log L_n}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log L_n}{\partial \theta_{k+r}} \end{vmatrix}$$

où $\underset{\sim}{\theta} = \underset{\sim}{\theta}(\underset{\sim}{U})$. Cette forme quadratique suit asymptotiquement la loi de χ^2 à k degrés de liberté et est la statistique proposée dans [45] pour $H\omega$. Maintenant, supposons qu'on remplace les $U_{k+1},\ldots,U_{k+r}$ par leurs estimateurs du maximum de vraisemblance sans restriction.

Considérons la surface $K_{c_n}(\underset{\sim}{\theta})$ définie par

$$(2.3.29)\quad \sum_{i=1}^{k}\sum_{j=1}^{k} U_i(\underset{\sim}{\theta})\,U_j(\underset{\sim}{\theta})\,a^*_{ij}(\underset{\sim}{\bar{\theta}}) = c_n \text{ où } c_n \to 0 \text{ quand } n \to \infty$$

$$\sum_{i=1}^{k+r} \gamma_{ti}(\underset{\sim}{\bar{\theta}})\,U_i(\underset{\sim}{\theta}) = \sum_{i=1}^{k+r} \gamma_{ti}(\underset{\sim}{\bar{\theta}})\,U_i(\underset{\sim}{\bar{\theta}})$$

pour $t = k+1,\ldots,k+r$

où $A^*(\underset{\sim}{\theta}) = [\,a^*_{ij}(\underset{\sim}{\theta})\,]$ et $\gamma_{ti}(\underset{\sim}{\theta})$ satisfont à la condition suivante : Il existe une matrice $[\beta_{ij}(\underset{\sim}{\theta})]$ i,j = 1,...,k telle que, si l'on construit la matrice $\mathcal{B}(\underset{\sim}{\theta})$ comme dans la proposition 2.3.2.

$$(2.3.30)\quad \mathcal{B}^T(\underset{\sim}{\theta})\;\Sigma^*(\underset{\sim}{\theta})\;\mathcal{B}(\underset{\sim}{\theta}) = I \text{ la matrice identique, où}$$

$$\Sigma^*(\underset{\sim}{\theta}) = \begin{bmatrix} A^*(\underset{\sim}{\theta}) & 0 \\ 0 & \Lambda^*_{22}(\underset{\sim}{\theta}) \end{bmatrix}$$

(analogue de $\Sigma(\underset{\sim}{\theta})$ donnée par (2.3.15)) .

Considérons les transformations

$$\theta'_i = \beta_{i1}(\underset{\sim}{\bar{\theta}})\, U_1(\underset{\sim}{\theta}) + \ldots + \beta_{ik}(\underset{\sim}{\bar{\theta}})\, U_k(\underset{\sim}{\theta})$$

$i = 1,\ldots,k$

$$\theta'_i = \gamma_{i1}(\underset{\sim}{\bar{\theta}})\, U_1(\underset{\sim}{\theta}) + \ldots + \gamma_{ik+r}(\underset{\sim}{\bar{\theta}})\, U_{k+r}(\underset{\sim}{\theta})$$

$i = k+1,\ldots,k+r$.

Avec ces transformations, la surface $K_{c_n}(\underset{\sim}{\bar{\theta}})$ se transforme en $K'_{c_n}(\underset{\sim}{\bar{\theta}})$ donnée par

$$(2.3.39') \qquad \sum_1^k \theta'^2_i = c_n$$

$$\theta'_t = \sum_{i=1}^{k+r} \gamma_{ti}(\underset{\sim}{\bar{\theta}})\, U_i(\underset{\sim}{\bar{\theta}})$$

pour $t = k+1,\ldots,k+r$.

Maintenant, des propositions 2.3.2. et 2.3.4. et de la discussion ci-dessus, on peut énoncer la proposition suivante

PROPOSITION 2.3.5. — *Soit* $\{W_n\}_{n \in \mathbf{N}^*}$ *une suite de régions critiques définies par* (2.3.27) *avec des estimateurs du maximum de vraisemblance sans restriction utilisés pour* $_2\underset{\sim}{U}$. *Soit* $K_{c_n}(\underset{\sim}{\bar{\theta}})$ *la famille des surfaces définie par* (2.3.29) . *Alors sous les hypothèses 1.1. - 1.6. et l'hypothèse* A , *pour tester l'hypothèse* $H\omega$ *donnée par* (2.3.26.), *la suite* $\{W_n\}_{n \in \mathbf{N}^*}$

(1) *a asymptotiquement la meilleure puissance constante sur la famille des surfaces* $K_{c_n}(\underset{\sim}{\bar{\theta}})$ *et*

(2) *est asymptotiquement la plus stricte.*

Remarque. — Il y a plusieurs méthodes disponibles pour les problèmes de test d'hypothèses concernant plusieurs paramètres. Le critère qu'on a considéré ci-dessus a l'avantage de ne pas nécessiter le calcul des estimateurs des paramètres quand on considère les hypothèses simples. Pour les hypothèses composées, on a un choix d'estimateurs pour les paramètres non spécifiés et tous les critères nous donnent des régions asymptotiquement semblables. Ceci n'exclut pas la possibilité que, si l'on remplace des paramètres non spécifiés par d'autres estimateurs, on peut encore se ramener aux propriétés asymptotiquement optimales.

Pour distinguer les différents critères avec les mêmes propriétés asymptotiques dans le cas d'échantillons finis, il semble qu'on doive comparer leur vitesse de convergence vers la loi asymptotique. Ceci demande la connaissance explicite de la forme de la loi de probabilité.

3. Une caractérisation de la convergence des procédures statistiques

3.0. Introduction.

Le problème de la convergence des procédures statistiques a été traité par plusieurs auteurs (voir par exemple Kraft [31] pour la bibliographie). Pour le problème du test d'une hypothèse simple contre une hypothèse alternative simple dans le cas de variables aléatoires indépendantes de même loi, Basu [12] a montré que le lemme de Neyman et Pearson donne un test convergent. Sur un espace produit infini

$$(\Omega, \mathcal{B}) = (\prod_{i=1}^{\infty} \mathcal{X}_i, \bigotimes_{1}^{\infty} \mathcal{S}_i) \quad \text{où} \quad \{(\mathcal{X}_i, \mathcal{S}_i)\}$$

est une suite d'espaces mesurables, Kakutani [28] a caractérisé la propriété d'orthogonalité de deux probabilités. Soit $\mathfrak{M}$ l'ensemble des mesures de probabilité sur $\mathcal{B}$, P et $Q \in \mathfrak{M}$ sont dits orthogonaux (et on note $P \perp Q$) s'il existe un ensemble $A \in \mathcal{B}$ tel que $P(A) = Q(\complement A) = 1$. Cette propriété est liée à la fonction $\rho : \mathfrak{M} \times \mathfrak{M} \to [0,1]$ définie par

$$\rho(P,Q) = \int_{\Omega} \sqrt{pq} \, d\mu$$

où p et q sont respectivement des densités de P et Q par rapport à une mesure σ-finie μ sur $\mathcal{B}$ en supposant que $P << \mu$ et $Q << \mu$ (Pour une famille finie ou dénombrable de mesures, une telle mesure existe toujours). Kraft [ibid] a fait une étude très étendue sur les propriétés de cette fonction en rapport avec les propriétés des procédures statistiques comme la convergence et la convergence uniforme des tests et des estimateurs du maximum de vraisemblance. Dans ce qui va suivre, en vue d'étudier la convergence d'un test et la convergence d'un estimateur, nous introduisons une fonction dans l'espace de probabilité sur un espace mesurable.

Nous en déduirons que, pour tester une hypothèse simple contre une hypothèse alternative simple, chaque fois qu'il existe un test convergent, le test donné par le lemme de Neyman-Pearson est un des tests les plus puissants, convergent (Most powerful consistent test). Le concept de convergence d'un estimateur a été examiné récemment par Rao [48] dans ses études systématiques sur les anomalies et irrégularités qui apparaissent dans la méthode du maximum de vraisemblance. Il a remarqué qu'un estimateur est convergent si la discrimination complète entre deux probabilités qui correspondent aux deux valeurs du paramètre au moyen d'un échantillon lorsque sa taille augmente indéfiniment, entraîne la discrimination

complète pour les lois asymptotiques correspondantes de cet estimateur. Avec la même fonction employée pour caractériser la convergence d'un test, nous donnerons une définition de la convergence d'un estimateur basée sur l'idée de Rao. Chaque fois qu'un estimateur est convergent en probabilité (ou presque sûrement) - convergence au sens classique - il est convergent au sens de la nouvelle définition. - convergence au sens large - Mais il existe des estimateurs qui ne sont pas convergents en probabilité et qui sont convergents au sens large. En outre, on peut observer que sous certaines hypothèses l'estimateur du maximum de vraisemblance est aussi convergent au sens large, ce qu'on verra dans le prochain chapitre. Ainsi, la classe des estimateurs dans laquelle on peut chercher un estimateur convergent est élargie. On fait quelques remarques concernant la nouvelle définition et la relation entre l'existence d'un test convergent et un estimateur convergent pour un paramètre.

3.1. Préliminaires.

Soit $\{(\mathcal{X}_i, \mathcal{S}_i)\}_{i \in \mathbf{N}^*}$ une suite d'espaces mesurables. Soient

$$(\Omega, \mathcal{B}) = \left(\prod_{i \in \mathbf{N}^*} \mathcal{X}_i , \bigotimes_{i \in \mathbf{N}^*} \mathcal{S}_i \right)$$

et

$$(\Omega^{(n)}, \mathcal{B}^{(n)}) = \left(\prod_{i=1}^{n} \mathcal{X}_i , \bigotimes_{i=1}^{n} \mathcal{S}_i \right) .$$

Pour chaque n, soit $P^{(n)}$ une mesure de probabilité sur $\mathcal{B}^{(n)}$ qui admet un prolongement unique P sur $\mathcal{B}$. D'ailleurs, si $\mathcal{S}$ est une sous-tribu de $\mathcal{B}$, la restriction de P sur $\mathcal{S}$ sera notée $P_{\mathcal{S}}$.

Soit P et Q deux mesures de probabilité sur $\mathcal{B}$ obtenues respectivement comme prolongement de $\{P^{(n)}\}$ et $\{Q^{(n)}\}$ sur $\{\mathcal{B}^{(n)}\}$. Supposons que, pour chaque n, il existe une partition mesurable de $\Omega^{(n)}$, $(A^{(n)}, \complement A^{(n)})$; telle que

$$(3.1.1.) \qquad P^{(n)}(A^{(n)}) = Q^{(n)}(\complement A^{(n)}) = \alpha_{A^{(n)}}(P, Q) .$$

Soit $\mathcal{F}^{(n)}$ l'ensemble de tous les éléments $A^{(n)} \in \mathcal{B}^{(n)}$ avec $P^{(n)}(A^{(n)}) = Q^{(n)}(\complement A^{(n)})$.

On pose

$$\alpha(P^{(n)}, Q^{(n)}) = \inf_{A^{(n)} \in \mathcal{F}^{(n)}} P^{(n)}(A^{(n)})$$

$$= \operatorname{Inf}_{A^{(n)} \in \mathcal{F}^{(n)}} Q(\complement A^{(n)}) \ .$$

Soit

$$(3.1.3.) \qquad d(P^{(n)}, Q^{(n)}) = \frac{1}{2} - \alpha(P^{(n)}, Q^{(n)})$$

$\alpha(P^{(n)}, Q^{(n)})$ est appelé l'erreur commune minimale pour la discrimination entre $P^{(n)}$ et $Q^{(n)}$ (ou la discrimination entre P et Q basée sur les n premières coordonnées d'un point observé de Ω) .

La fonction $d(P^{(n)}, Q^{(n)})$ satisfait aux propriétés suivantes

(1) $\quad d(P^{(n)}, Q^{(n)}) = d(Q^{(n)}, P^{(n)})$

(2) $\quad 0 \leqslant d(P^{(n)}, Q^{(n)}) \leqslant \frac{1}{2} \qquad d(P^{(n)}, Q^{(n)}) = 0$

si $P^{(n)} \equiv Q^{(n)}$ et $d(P^{(n)}, Q^{(n)}) = \frac{1}{2}$, si $P^{(n)} \perp Q^{(n)}$

(3) si P_1, P_2, P_3 sont trois mesures de probabilité sur $\mathcal{B}$ obtenues respectivement comme prolongement unique de $\{ P_1^{(n)} \}$, $\{ P_2^{(n)} \}$, et $\{ P_2^{(n)} \}$ sur $\{ \mathcal{B}^{(n)} \}$ on a

$$d(P_1^{(n)}, P_2^{(n)}) + d(P_2^{(n)}, P_3^{(n)}) \geqslant d(P_1^{(n)}, P_3^{(n)})$$

c'est-à-dire $d(P^{(n)}, Q^{(n)})$ satisfait à l'inégalité triangulaire (4). Soient $p^{(n)}$ et $q^{(n)}$ des densités de $P^{(n)}$ et $Q^{(n)}$ par rapport à une mesure σ-finie par rapport à laquelle $P^{(n)}$ et $Q^{(n)}$ sont absolument continues. La partition mesurable donnée par

$$A^{(n)} = \{ \omega^{(n)} \mid q^{(n)}(\omega^{(n)}) \geqslant k\, p^{(n)}(\omega^{(n)}) \}$$

où $k \geqslant 0$ est choisi telle que $P^{(n)}(A^{(n)}) = Q^{(n)}(\complement A^{(n)})$, donne $\alpha(P^{(n)}, Q^{(n)}) = P^{(n)}(A^{(n)}) = Q^{(n)}(\complement A^{(n)})$ (par exemple [1]) .

LEMME 3.1.1. – *Soient* P et Q *deux mesures de probabilité sur* $\mathcal{B}$. *Soit* $\mathcal{D}$ *une sous-tribu de* $\mathcal{B}$. *Alors*

$$\alpha(P, Q) \leqslant \alpha(P_{\mathcal{D}}, Q_{\mathcal{D}}) .$$

DÉMONSTRATION. – Soit $\mathcal{F}$ la classe des ensembles A de $\mathcal{B}$ telle que $P(A) = Q(\complement A)$

$$\alpha(P, Q) = \inf_{A \in \mathcal{F}} P(A) = \inf_{A \in \mathcal{F}} Q(\complement A) .$$

On a

$$\alpha(P_{\mathcal{D}}, Q_{\mathcal{D}}) = \inf_{A \in \mathcal{F} \cap \mathcal{D}} P_{\mathcal{D}}(A) \geqslant$$

$$\geqslant \inf_{A \in \mathcal{F}} P(A) = \alpha(P, Q) ,$$

ce qui démontre le lemme.

En particulier, pour les mesures $\{P^{(n)}\}$ et $\{Q^{(n)}\}$ de probabilité sur $\mathcal{B}^{(n)}$ et leur prolongements uniques P et Q sur $\mathcal{B}$ on a

(3.1.5.) $$\alpha(P^{(n)}, Q^{(n)}) \geqslant \alpha(P^{(n+1)} Q^{(n+1)}) \geqslant \ldots \geqslant \alpha(P, Q)$$

et par conséquent

(3.1.6.) $$d(P^{(n)}, Q^{(n)}) \leqslant d(P^{(n+1)}, Q^{(n+1)}) \leqslant \ldots \leqslant d(P, Q)$$

Donc

$\{ d(P^{(n)}, Q^{(n)}) \}_{n \in \mathbf{N}^*}$ est une suite croissante.

3.2. Tests convergents

Dans ce paragraphe, nous considérons le problème de l'existence des tests convergents (1) pour les hypothèses simples contre les hypothèses alternatives simples et (2) les hypothèses composées.

PROPOSITION 3.2.1. — *Les trois énoncés suivants sont équivalents*

(i) $d(P^{(n)}, Q^{(n)}) \uparrow \frac{1}{2}$

(ii) *Il existe un test convergent pour l'hypothèse* H_o : *la vraie probabilité est* P *contre* H_1 : *elle est* Q

(iii) $P \perp Q$.

DÉMONSTRATION. — Pour simplifier l'écriture, nous posons
$d^{(n)} = d(P^{(n)}, Q^{(n)})$ et $\alpha^{(n)} = \alpha(P^{(n)}, Q^{(n)})$

(i) $\Longrightarrow$ (ii). En effet, pour chaque n , on a
$\alpha^{(n)} = \alpha(P^{(n)}, Q^{(n)}) = P^{(n)}(A^{(n)})$

où

(3.2.1.) $A^{(n)} = \{ (x_1,\ldots,x_n) \mid q^{(n)}(x_1,\ldots,x_n) \geqslant k_n p^{(n)}(x_1,\ldots,x_n) \}$,
où k_n est choisi de telle sorte que $P^{(n)}(A^{(n)}) = Q^{(n)}(\complement A^{(n)})$.

Définissons alors

(3.2.2.) $\varphi_n(x_1,\ldots,x_n) = 1$ si $(x_1,\ldots,x_n) \in A^{(n)}$
$= 0$ ailleurs

Alors
$$E_P(\varphi_n) = E_Q(1 - \varphi_n) = \alpha^{(n)} .$$

Mais $d^{(n)} = \frac{1}{2} - \alpha^{(n)}$. Donc $d^{(n)} \uparrow \frac{1}{2}$ entraîne que $\alpha^{(n)} \downarrow 0$ et par conséquent, φ_n est un test convergent pour H_o contre H_1 .

(ii) $\Longrightarrow$ (iii) (Démonstration due à Kraft) : En effet, soit $\{\varphi_n\}_{n \in \mathbf{N}^*}$ un test convergent de H_o contre H_1 . Alors $E_P(\varphi_n) \downarrow 0$ et $E_Q(\varphi_n) \uparrow 1$. Choisissons une sous-suite $\{\varphi_{n_k}\}_{k \in \mathbf{N}^*}$ telle que

$$E_P(\varphi_{n_k}) \leqslant \varepsilon_k \quad \text{et} \quad E_Q(\varphi_{n_k}) \geqslant 1 - \varepsilon_k$$

où $\{\varepsilon_k\}_{k \in \mathbf{N}^*}$ est une suite de nombres positifs tels que $\Sigma \varepsilon_k < \infty$.

Posons encore $\varphi(\omega) = \limsup_{k \to \infty} (\varphi_{n_k}(\omega)) \quad \forall\, \omega \in \Omega$

Alors $E_P(\varphi) = \lim_{i \to \infty} E_P(\sup_{k \geqslant i} \varphi_{n_k}) \leqslant$

$$\leqslant \lim_{i \to \infty} \sum_{k \geqslant i}^{\infty} E_P(\varphi_{n_k}) = 0 \ .$$

Puisque $0 \leqslant \varphi \leqslant 1$, il résulte que $P\{\varphi(\omega) = 0\} = 1$. De la même façon, on voit que $Q\{\varphi(\omega) = 0\} = 0$ de sorte que $P \perp Q$.

Pour montrer que (iii) $\Longrightarrow$ (i), nous remarquons que

$$\alpha^{(n)} = P(A^{(n)}) = \int_{A^{(n)}} p^{(n)} d\mu = \int_{\complement A^{(n)}} q^{(n)} d\mu$$

où $p^{(n)} = E\mu\,(p \mid \mathcal{B}^{(n)})$ et $q^{(n)} = E_\mu(q \mid \mathcal{B}^{(n)})$.

Puisque $p^{(n)}$ et $q^{(n)}$ sont des martingales (Doob, ch. V)
$\mu\,(p^{(n)} \to p) = 1$ et $\mu\,(q^{(n)} \longrightarrow q) = 1$ (μ est aussi une probabilité).
Aussi $P(A^{(n)}) \downarrow 0$ entraîne que $\alpha^{(n)} \downarrow 0$ et par conséquent $d^{(n)} \uparrow \frac{1}{2}$, ce qui établit l'équivalence des trois énoncés.

Au cours de la démonstration, nous avons établi aussi le

COROLLAIRE. — Chaque fois qu'il existe un test convergent pour H_0 *contre* H_1 , *le lemme de Neyman-Pearson fournit un des tests les plus puissants et convergents.*

Ceci est une conséquence de l'équivalence entre (i) et (ii) .

Exemple. — Des résultats précédents, on peut déduire que si $\{X_{n,\, n \in \mathbf{N}^*}\}$ est un processus stochastique ergodique strictement stationnaire avec des lois P sous l'hypothèse nulle H_0 et Q sous l'hypothèse H_1 , un des tests les plus puissants, convergent, est fourni par le lemme de Neyman-Pearson, pourvu que P ne soit pas absolument continu par rapport à Q .

On a $\alpha^{(n)} = P(A^{(n)})$ où

$$A^{(n)} = \left\{ (x_1,\dots,x_n) \;\middle|\; \frac{q^{(n)}(x_1,\dots,x_n)}{p^{(n)}(x_1,\dots,x_n)} \geqslant k_n \right\} \times \prod_{n+1}^{\infty} \mathcal{X}_i$$

$k_n \geqslant 0$ étant choisi tel que $P(A^{(n)}) = Q(\complement A^{(n)})$. On sait que $A^{(n)}$ est une région critique la plus puissante pour H_o contre H_1 .

Or $$\frac{q^{(n)}(x_1,\dots,x_n)}{p^{(n)}(x_1,\dots,x_n)} \longrightarrow 0, \; P \text{ presque sûrement}$$

pourvu que P ne soit pas absolument continue par rapport à Q (Brieman [16]) .

Donc $k_n \longrightarrow 0$ quand $n \to \infty$.

Or $$Q(A^{(n)}) = Q\left[\left\{ \frac{q^{(n)}(x_1,\dots,x_n)}{p^{(n)}(x_1,\dots,x_n)} \geqslant k_n \right\} \times \prod_{n+1}^{\infty} \mathcal{X}_i \right]$$

$$= Q\left[\left\{ \frac{p^{(n)}(x_1,\dots,x_n)}{q^{(n)}(x_1,\dots,x_n)} \leqslant \frac{1}{k_n} \right\} \times \prod_{n+1}^{\infty} \mathcal{X}_i \right]$$

qui tend vers 1 , ce qui donne le résultat.

De cet exemple, on peut déduire que le lemme de Neyman-Pearson fournit un des tests les plus puissants, convergent, pour une chaîne de Markov ou même pour un processus de Markov à temps continu sous certaines hypothèses de régularité (voir ch.IV) toujours pourvu que la probabilité sous H_o ne soit pas absolument continue par rapport à la probabilité sous H_1 .

3.2.(2). Discrimination.

La fonction $d(P,Q)$ est introduite par Rao [45] pour étudier les problèmes de discrimination entre un nombre fini de mesures de probabilité. Supposons que T soit une application mesurable de $(\Omega, \mathcal{B})$ sur

$(\mathcal{Y}, \mathcal{C})$. Supposons qu'il existe une partition mesurable de $\mathcal{Y}$, $A_T \cup \complement A_T = \mathcal{Y}$ telle que

$$P_T(A_T) = Q_T(\complement A_T) = \alpha_T(P_T, Q_T) ,$$

erreur commune minimale dans l'espace $(\mathcal{Y}, \mathcal{C})$. Soit $\alpha(P,Q)$ l'erreur commune minimale dans $(\Omega, \mathcal{B})$; considérons la partition $(T^{-1}(A_T)$, $T^{-1}(\complement A_T))$ de Ω . On a manifestement

$$P(T^{-1}A_T) = Q(T^{-1}(\complement A_T)) = \alpha_T ,$$

c'est-à-dire que, pour cette partition, les deux erreurs de discrimination sont égales. Comme α est la valeur minimale de cette erreur, on a $\alpha(P,Q) \leqslant \alpha_T$ et par conséquent $d(P,Q) \geqslant d(P_T, Q_T)$, c'est-à-dire que l'erreur de discrimination ne diminue pas si l'on emploie une statistique au lieu de l'échantillon (ceci résulte aussi du lemme 3.1.1.). Mais si T est une statistique exhaustive pour la famille de mesures, alors

$$d(P, Q) = d(P_T, Q_T) \quad \text{(voir par exemple [2]).}$$

Ceci implique que l'on peut faire une discrimination entre les mesures en utilisant une statistique exhaustive qui sera aussi bonne que la discrimination faite en utilisant tout l'échantillon (Théorème de Rao-Blackwell) .

3.2.(3), Maintenant considérons le cas de l'hypothèse composée
$H_o : \theta \in \Theta_o$ $\qquad$ $H_1 : \gamma \in \Theta_1$
Θ_o et Θ_1 étant deux boréliens de $\mathbf{R}^k$.

Lorsque l'hypothèse n'est pas simple, il n'est plus possible de caractériser l'existence d'un test convergent par la fonction $d(P,Q)$ définie seulement entre les éléments de H_o et H_1 . Supposons que Θ_o et Θ_1 sont munis respectivement des tribus $\mathcal{A}_o$ et $\mathcal{A}_1$, traces de la tribu borélienne de $\mathbf{R}^k$. Nous supposons que $P(. ; \theta)$, $Q(. ; \gamma)$ sont respectivement mesurables $\mathcal{B} \otimes \mathcal{A}_o$ et $\mathcal{B} \otimes \mathcal{A}_1$. Soient λ et ν , deux probabilités respectivement sur $\mathcal{A}_o$ et $\mathcal{A}_1$. Considérons les mesures de probabilité

$$R_\lambda(A) \quad \text{et} \quad S_\nu(A) \quad \text{sur} \quad \mathcal{B} \quad \text{définie par}$$

(3.2.8.) $$R_\lambda(A) = E_\lambda(P(A;\theta)) = \int_\Theta P(A;\theta)\,\lambda(d\theta)$$

$$S_\nu(A) = E_\nu(Q(A;\gamma)) = \int_\Theta Q(A;\gamma)\,\nu(d\gamma)$$

$\forall A \in \mathcal{B}$.

Considérons les classes de mesures $\{R_\lambda\}$ et $\{S_\nu\}$ obtenues pour toutes les mesures de probabilité λ et ν respectivement sur $\mathcal{C}_0$ et $\mathcal{C}_1$. Il est évident qu'elles contiennent les classes $\{P(.;\theta)\ \theta \in \Theta_0\}$ et $\{Q(.;\gamma)\ , \gamma \in \Theta_1\}$ respectivement. Si les éléments de ces dernières classes sont absolument continues par rapport à une mesure σ-finie μ , les éléments des classes $\{R_\lambda\}$ et $\{S_\nu\}$ sont aussi absolument continues par rapport à μ . Il est possible de caractériser l'existence d'un test convergent entre les éléments des nouvelles classes.

PROPOSITION 3.2.2. – *Les trois énoncés suivants sont équivalents*

(i) $$d(E_\lambda(P^{(n)}(.;\theta)),\ E_\nu(Q^{(n)}(.;\nu))) \uparrow \frac{1}{2}$$

(ii) *Il existe un test convergent pour* λ*-presque tout* $\theta \in \Theta_0$ *contre* ν*-presque* $\gamma \in \Theta_1$

(iii) $$E_\lambda(P(.;\theta)) \quad E_\nu(Q(.;\gamma))\ .$$

DÉMONSTRATION. – (i) $\Longrightarrow$ (ii) Pour n fini, considérons la région

$$A^{(n)} = \left\{(x_1,\ldots,x_n) \;\middle|\; \frac{s(x_1,\ldots,x_n)}{r(x_1,\ldots,x_n)} \geqslant k_n \right\} \times \prod_{n+1}^{\infty} \mathcal{X}_i$$

telle que $R_\lambda(A^{(n)}) = S_\nu(\complement A^{(n)})$ qui sera l'erreur commune minimale pour la discrimination entre R_λ et S_ν .

$$R_\lambda(A^{(n)}) = \int_{A^{(n)}} \int_{\Theta_0} p(\omega;\theta)\,d\lambda\,d\mu$$

(3.2.10)

$$= \int_{\Theta_o} \left(\int_{A^{(n)}} p(\omega ; \theta)\, d\mu \right) d\lambda$$

d'après le théorème de Fubini.

Donc $R_\lambda (A^{(n)}) \to 0 \Longrightarrow \int_{A^{(n)}} p(\omega ; \theta)\, d\mu \to 0$

en probabilité - λ quand $n \to \infty$.

De la même façon

$$S_\nu (\complement A^{(n)}) \to 0 \longrightarrow \int_{\complement A^{(n)}} q(\omega ; \gamma)\, d\mu \to 0$$

en probabilité - ν quand $n \to \infty$.

Donc, il existe une sous-suite $\{A^{(n_k)}\}_{k \in \mathbf{N}^*}$ telle que

$$\int_{A^{(n_k)}} p(\omega ; \theta)\, d\mu \to 0 \qquad \lambda \text{ - presque sûrement}$$

(3.2.11.)

$$\int_{A^{(n_k)}} q(\omega ; \gamma)\, d\mu \to 0 \qquad \nu \text{ - presque sûrement}$$

quand $n \to \infty$.

Donc, il existe un test convergent pour λ - presque tout $\theta \in \Theta_o$ contre ν - presque tout $\gamma \in \Theta_1$.

(ii) $\Longrightarrow$ (iii) Considérons un test convergent $\{\varphi_n\}$ pour $P(. ; \theta)$ $\theta \in \Theta_o$ contre $Q(. ; \gamma)$, $\gamma \in \Theta_1$; ce qui existe par (ii) .

Alors, il en résulte que

$$\int_{\Omega} \varphi_n \, dP(.\,;\theta) \to 0 \quad \lambda\text{-p.p}$$
$$\int_{\Omega} \varphi_n \, dQ(.\,;\gamma) \to 1 \quad \nu\text{-p.p} \qquad \text{lorsque } n \to \infty$$

Par conséquent,

$$\int_{\Theta_o} \int_{\Omega} \varphi_n \, dP(.\,;\theta) \, d\lambda \to 0 \quad \text{et}$$

$$\int_{\Theta_1} \int_{\Omega} \varphi_n \, dQ(.\,;\gamma) \, d\nu \to 1$$

De plus, par une application du théorème de Fubini

$$\int_{\Omega} \varphi_n \, d\mu \int_{\Theta_o} p(x\,;\theta) \, d\lambda \to 0$$
$$\int_{\Omega} \varphi_n \, d\mu \int_{\Theta_1} q(x\,;\gamma) \, d\nu \to 1 \qquad \text{quand } n \to \infty .$$

Donc, comme dans la proposition précédente, il résulte que
$E_\lambda (P(.\,;\theta))$ $E_\nu (R(.\,;\gamma))$

(iii) $\Longrightarrow$ (i) se démontre comme dans la proposition 3.2.1.

Il est clair que $d(P^{(n)}, Q^{(n)}) \uparrow \frac{1}{2}$ est une condition nécessaire pour l'existence d'un test convergent dans le cas de l'hypothèse composée. Kraft a cité un exemple dû à Hodges [ibid] , qui montre que la condition n'est pas suffisante en général. Le théorème suivant de Kraft donne une condition nécessaire et suffisante pour qu'il existe un test convergent lorsqu'il s'agit d'hypothèse composée.

THÉORÈME DE KRAFT. — Pour qu'il existe un test convergent pour une hypothèse nulle composée dénombrable contre une hypothèse alternative dénombrable, il faut et il suffit qu'il existe un test convergent pour chaque paire d'hypothèses simples définies par $\theta \in \Theta_o$ *et* $\gamma \in \Theta_1$.

3.3. Convergence des estimateurs

Il y a deux types de convergence des estimateurs couramment utilisés. Convergence en probabilité (P C) (et parfois convergence presque sûre) et convergence au sens de Fisher. La première est la plus connue et la plus appliquée. En général, on a pensé que les estimateurs du maximum de vraisemblance sont PC. Mais il existe plusieurs cas (voir [47] pour les références) où des estimateurs du maximum de vraisemblance ne sont pas PC. Tout ceci a conduit à un nouvel examen du concept de convergence. Rao [48] dans ses études systématiques des concepts de convergence et de l'efficacité asymptotique des estimateurs, a remarqué qu'un estimateur est dit convergent, si l'on peut effectuer la discrimination complète entre les deux mesures induites par l'estimateur, chaque fois que les deux mesures qui correspondent à l'échantillon sont orthogonales lorsque la taille de l'échantillon augmente indéfiniment. Dans ce cas, l'estimateur est dit convergent au sens large. Dans ce cadre, soit $\{P(.\,;\theta)\ \theta \in \Theta\}$ une famille de mesures de probabilités définies comme ci-dessus, dépendans d'un paramètre $\theta \in \Theta$. Soit $T_n(\underset{\sim}{X}^{(n)})$ un estimateur de θ basé sur un échantillon de taille n noté $\underset{\sim}{X}^{(n)}$. Soit $\varphi_{T_n}(.\,;\underset{\sim}{\theta})$ la probabilité induite par T_n.

DÉFINITION 3.3.1. — Un estimateur T_n est dit convergent au sens large si $d(\varphi_{T_n}(.\,;\theta_1), \varphi_{T_n}(.\,;\theta_2))$ tend vers $\frac{1}{2}$ chaque fois que $d(P^{(n)}(.\,;\theta_1), P^{(n)}(.\,;\theta_2)) \to \frac{1}{2}$ pour $\theta_1 \neq \theta_2$. Si la convergence est uniforme en θ_1 et θ_2, l'estimateur est dit uniformément convergent au sens large.

Avec cette définition nous allons faire quelques observations qui sont immédiates après la définition. Tout d'abord, remarquons qu'avec cette définition, on n'exige pas que $T_n \to \theta$, en probabilité. En effet, il n'est pas nécessaire que T_n tende en probabilité vers une fonction $g(\theta)$ de θ quand $n \to \infty$.

Même s'il existe une loi asymptotique pour T_n, cette définition est applicable. Mais si un estimateur est convergent en probabilité (ou preque sûrement) c'est-à-dire $T_n \to g(\theta)$ en probabilité. $g(\theta)$ étant une bijection de Θ dans $\mathbf{R}$, alors T_n est convergent au sens large. Inversement, si T_n est convergent au sens large, T_n est convergent en probabilité, seulement si de plus la loi asymptotique de T_n est dégénérée pour chaque θ.

Remarque 1. — Avec cette définition, certains estimateurs qui ne sont pas convergents en probabilité, sont convergents au sens large.

Remarque 2. — Cette définition s'applique directement au cas du paramètre vectoriel.

Remarque 3. — Supposons qu'il existe un test convergent pour toute hypothèse simple contre toute hypothèse alternative simple pour $\theta \in \Theta$. Supposons qussi qu'il existe une statistique exhaustive pour la famille des probabilités. Alors la statistique exhaustique est convergente au sens large puisqu'un test basé sur la statistique exhaustive est aussi puissant que tous les autres tests.

Remarque 4. — S'il existe un estimateur convergent au sens large, puisque

$$d(\varphi_{T_n}(.\,;\theta_1), \varphi_{T_n}(.\,;\theta_2) \leqslant d(P^{(n)}(.\,;\theta_1), P^{(n)}(.\,;\theta_2))\,,$$

il en résulte qu'il existe un test convergent pour θ_1 contre θ_2.

Remarque 5. — Chaque fois qu'il existe un estimateur convergent en probabilité pour θ, il existe aussi un test convergent pour θ contre toute hypothèse alternative simple définie par une valeur différente de θ.

4. Estimation pour des processus stochastiques

4.0. Introduction

Dans ce chapitre, comme dans le suivant, on va considérer le problème de l'efficacité des estimateurs. Parmi plusieurs idées fondamentales, le concept classique de l'efficacité d'un estimateur est dû à R.A. Fisher. Fisher a rattaché la variance asymptotique d'un estimateur à l' "information à la limite (limiting information)" dans l'échantillon. Il existe plusieurs études très étendues sur l'efficacité d'un estimateur et en particulier l'efficacité d'un estimateur du maximum de vraisemblance. (voir Lecam [32] Rao [47] pour la bibliographie). Ces études ont aussi montré quelques anomalies et irrégularités dans la méthode du maximum de vraisemblance. Ceci a conduit à examiner de nouveau le concept de l'efficacité et de la méthode du maximum de vraisemblance. Celui-ci a été traité par plusieurs auteurs. Dans une séries d'articles, Rao [47] , [48] , a étudié systématiquement les concepts de convergence et d'efficacité d'un estimateur ; il a donné une nouvelle définition de l'efficacité et a réexaminé la méthode du maximum de vraisemblance. Ici, nous formulons dans le cas général une nouvelle définition dont nous justifierons la nécessité. On démontre que la méthode du maximum de vraisemblance donne des estimateurs asymptotiquement efficaces. On démontre aussi que dans les cas particuliers, on peut à la fois affaiblir les hypothèses et obtenir des résultats plus précis. En outre, on verra que la nouvelle définition dont nous avons parlé ci-dessus, implique les exigences de Wald [55] dans sa recherche pour une nouvelle définition de l'efficacité, tandis qu'une variance implique celles de Bahadur [8] .

Cramer-Darmois-Fréchet-Rao ont obtenu, sous les conditions de régularité une "borne inférieure" pour la variance d'un estimateur qui ne dépend pas de la méthode d'estimation. Chapman et Robbins [18] ont affaibli les hypothèses de dérivabilité faites par les auteurs cités ci-dessus et ont démontré un résultat de même type. Par conséquent, la variance minimale d'un estimateur est exigée comme une propriété optimale avec laquelle un estimateur est dit efficace. Doob, Dugué, Hotelling parmi les autres ont étudié les méthodes d'estimation satisfaisant à cette propriété optimale et ont notamment démontré que la méthode du maximum de vraisemblance donne des estimateurs asymptotiquement efficaces. Cependant, Hodges (voir Lecam [32]) a construit l'exemple ci-dessous qui a fait douter de l'intérêt de ce critère.

Considérons une variable aléatoire X qui suit une loi gaussienne d'espérance mathématique θ inconnue et de variance unité. L'estimateur du maximum de vraisemblance obtenu à partir d'un échantillon $(X_1,\ldots,X_n)$ de taille n est la moyenne de l'échantillon $\bar{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$. $\bar{X}_n$ a la variance égale à la borne inférieure de la variance. Mais considérons un estimateur T_n défini par

$$T_n = \bar{X}_n \quad \text{si} \quad |\bar{X}_n| > n^{-1/4}$$

$$= \alpha \bar{X}_n \quad \text{si} \quad |\bar{X}_n| \leqslant n^{-1/4}$$

où α est une constante arbitraire. La variance asymptotique de T_n est donnée par

$$V(T_n) = 1 \quad \text{si} \quad \theta \neq 0$$

$$= \alpha^2 \quad \text{si} \quad \theta = 0$$

et peut être prise ainsi arbitrairement petite pour la valeur $\theta = 0$ du paramètre. Lecam [ibid] a généralisé cet exemple et il démontre qu'il est possible de construire des estimateurs dont la variance est inférieure à la valeur considérée par les auteurs cités ci-dessus comme "la borne inférieure de la variance" (estimateur super-efficace) sur un ensemble de mesure de Lebesgue nulle et cet ensemble ne peut pas être un intervalle. Bahadur [9] a montré qu'en général, cet ensemble ne peut pas être élargi. Depuis il y a eu plusieurs tentatives pour étudier la signification exacte de ce critère d'optimalité. Kallianpur et Rao [30] ont obtenu des conditions qui garantissent que la variance d'un estimateur satisfaisant à ces conditions ne peut pas être inférieure à la borne de Cramer.Rao. Plus tard, Kallianpur [29] a affaibli les hypothèses et démontré que dans quelques cas, les estimateurs du maximum de vraisemblance sont asymptotiquement efficaces. De son côté, Rao [49] a démontré que sous certaines hypothèses de convergence uniforme, le critère de variance minimale est encore valable.

Cependant, même s'il n'existe pas une borne inférieure pour la variance asymptotique, la question non résolue est ; dans l'exemple ci-dessus, doit-on préférer T_n à $\bar{X}_n$ parce que sa variance est plus petite que celle de $\bar{X}_n$ pour une certaine valeur du paramètre ? Peut-on éviter tout ceci si l'on considère seulement les estimateurs efficaces au sens de la variance minimale et ne pas essayer d'améliorer la variance d'un estimateur pour certaines valeurs du paramètre ? Considérons l'exemple sui-

vant donné par Rao [49] pour le même cas gaussien d'espérance mathématique θ et de variance unité.

Considérons un estimateur

$$W_n : \bar{X}_n \quad \text{si} \quad |\bar{X}_n| > n^{-1/4} =$$

$$= \left(\frac{2}{\pi}\right)^{1/2} X_{(m)} \quad \text{si} \quad |\bar{X}_n| \leqslant n^{-1/4}$$

où $X_{(m)}$ est la médiane de l'échantillon. W_n a la même variance asymptotique que $\bar{X}_n$ pour tout θ, est convergent, et est asymptotiquement gaussien comme $\bar{X}_n$. Comment peut-on distinguer W_n de $\bar{X}_n$ à partir des propriétés classiques d'optimalité ? Cependant pour n assez grand, $\bar{X}_n$ est plus concentré autour de la vraie valeur de θ que W_n. Un estimateur ayant plus de concentration que l'autre pour tout θ est plus utile pour les tests d'hypothèses basés sur les estimateurs concernant le paramètre. Si l'on fait seulement une comparaison des variances asymptotiques, on ne peut pas apprécier cet aspect d'un estimateur comme le montre cet exemple. Mais si un estimateur n'a pas une concentration uniformément meilleure que l'autre, il devient difficile de choisir entre les deux sans utiliser d'autres critères. Par exemple T_n a une meilleure concentration que $\bar{X}_n$ autour de $\theta = 0$, mais pour les autres valeurs de θ, c'est l'inverse qui se produit. Bahadur [8] a considéré la concentration d'un estimateur autour de la vraie valeur dans les intervalles fixes quand $n \to \infty$ comme critère d'optimalité d'un estimateur. Pour faire une comparaison entre T_n, $\bar{X}_n$ et W_n on peut examiner leur rôle pour tester des hypothèses concernant le paramètre. Avec cette idée, Rao [47], [48] a donné une nouvelle définition de l'efficacité asymptotique dans le cas de variables aléatoires indépendantes.

Suivant la même idée générale, dans ce qui va suivre, nous formulons dans le cas général une définition de l'efficacité asymptotique qui, d'ailleurs, entraîne celle donnée par Wald [55]. Une version plus stricte de cette définition implique celle de Bahadur [8] dans le cas de variables aléatoires indépendantes. On peut aussi observer que, dans le cas classique, on exige qu'un estimateur suive une loi gaussienne. Ceci est non seulement superflu, mais n'est peut-être pas toujours vérifié dans le cas des processus stochastiques généraux. La définition donnée ici ne fait aucune hypothèse de cette sorte.

4. 1. Préliminaires.

Soit $(\mathcal{X}, \mathcal{S})$ un espace polonais muni de sa tribu $\mathcal{S}$ d'ensembles boréliens. Soit $(\Omega, \mathcal{B}) = (\mathcal{X}^{\mathbf{N}^*}, \mathcal{S}^{\mathbf{N}^*})$. Soit $\{X_n, n \in \mathbf{N}^*\}$ un processus stochastique à temps discret défini sur $(\Omega, \mathcal{B})$ à valeurs dans $\mathcal{X}$.

Dans le cas pratique, chaque point $x \in \mathcal{X}$ peut être considéré comme un état d'un système afin que pour chaque $\omega \in \Omega$, $X_n(\omega)$ regardé comme une fonction de $n \in \mathbf{N}^*$, donne les états successifs du système. Ceci est appelé une réalisation du processus (ou une trajectoire). On suppose que le système est gouverné par une probabilité $P(.\,, \theta)$ sur $\mathcal{B}$ obtenue comme prolongement unique des probabilités définies sur les produits finis $\{(\mathcal{X}^{(n)}, \mathcal{S}^{(n)}\}_{n \in \mathbf{N}^*}$ satisfaisant aux conditions de compatibilité.

L'existence d'une telle mesure est connue (cf. [38] ch. III).

On suppose que la forme de la loi est connue mais dépend d'un paramètre $\theta \in \Theta$, ouvert de $\mathbf{R}$.

On peut supposer que $P(.\,; \theta)$ est la loi qui gouverne l'évolution du système. On a une famille de mesures $\{P(.\,; \theta)\,, \theta \in \Theta\}$ sur $\mathcal{B}$ et l'on veut faire une estimation, de la vraie valeur du paramètre, basée sur un échantillon du processus - c'est-à-dire un morceau de la trajectoire $\{X_n, n \in \mathbf{N}^*\}$. Nous ne nous occuperons que des processus stochastiques strictement stationnaires.

Donc, les probabilités $P_{n_1, \ldots, n_m}(.\,; \theta)$ sur $\mathcal{S}^{(n)}$ peuvent se noter $P^{(m)}(.\,; \theta)$ $\forall m \in \mathbf{N}^*$ sans préciser les coordonnées explicitement.

On fait les hypothèses suivantes concernant les probabilités

Hypothèse 4.1. – Pour que la fonction de vraisemblance existe, supposons qu'il existe, pour chaque n , une mesure σ-finie $\lambda^{(n)}$ sur $\mathcal{S}^{(n)}$ telle que $\forall \theta \in \Theta$, $P^{(n)}(.\,; \theta) \ll \lambda^{(n)}$ c'est-à-dire

$$(4.1.1.) \quad P^{(n)}(A^{(n)}; \theta) = \int_{A^{(n)}} f_n(x_1, \ldots, x_n; \theta)\, d\lambda^{(n)}, \forall A^{(n)} \in \mathcal{S}^{(n)}$$

Il s'ensuit que, si $T_n(X_1,\ldots,X_n)$ est une statistique et $\Phi_{T_n}(.\,;\theta)$ la mesure induite par T_n, alors $\Phi_{T_n}(.\,;\theta) \ll \lambda'$ la mesure image de $\lambda^{(n)}$ par T_n. Soit $\varphi_{T_n}(X\,;\theta)$ sa densité.

Hypothèse 4.2. — Pour chaque n il existe un borélien $D^{(n)} \in \mathcal{S}^{(n)}$ indépendant de $\theta \in \Theta$, tel que $f_n(X_1,\ldots,X_n\,;\theta) > 0$, $\lambda^{(n)}$ presque sûrement et $P^{(n)}(D^{(n)}\,;\theta) = 1$, $\forall\,\theta \in \Theta$.

Hypothèse 4.3. — Les dérivées partielles $\dfrac{\partial^i f_n}{\partial\theta^i}$, $(i = 1, 2)$ existent pour tout $\theta \in \Theta$ $\lambda^{(n)}$- presque partout et sont continues en $\theta \in \Theta$. De plus, pour chaque $A^{(n)} \in \mathcal{S}^{(n)}$

$$(4.1.2.)\qquad \frac{\partial^i}{\partial\theta^i} \int_{A^{(n)}} f_n \, d\lambda^{(n)} = \int_{A^{(n)}} \frac{\partial^i f_n}{\partial\theta^i} \, d\lambda^{(n)}$$

pour $i = 1, 2$.

Et pour $\forall\, A \in \mathcal{A}$

$$\frac{\partial}{\partial\theta} \int_A \varphi_{T_n} \, d\lambda' = \int_A \frac{\partial \varphi_{T_n}}{\partial\theta} \, d\lambda' .$$

L'hypothèse 4.3. implique que $E_\theta\left(\dfrac{\partial \log f_n}{\partial\theta}\right) = 0$ et

$$E_\theta\left\{\left(\frac{\partial \log f_n}{\partial\theta}\right)^2\right\} = -\,E\left\{\frac{\partial^2 \log f_n}{\partial\theta^2}\right\} \quad \text{notée} \quad I_n(\theta),$$

information de Fisher dans l'échantillon.

Hypothèse 4.4. — $I_n(\theta)$ est continu en θ et

$$(4.1.3.)\qquad \lim_{n \to \infty} I_n(\theta) = \infty \,, \quad \forall\, \theta \in \Theta .$$

De plus, on peut supposer que $I_n(\theta) > 0$, $\forall n \in \mathbf{N}^*$, et $\forall\,\theta \in \Theta$.

Nous utilisons les notations suivantes ; soit

$I_{T_n}(\theta) = E_\theta \left[\left(\frac{\partial \log \varphi_n}{\partial \theta} \right)^2 \right]$ que nous supposons, sans nuire à la généralité, positive $\forall$ n .

$$(4.1.4.) \qquad Z_n(\theta) = [I_n(\theta)]^{-\frac{1}{2}} \left(\frac{\partial \log f_n}{\partial \theta} \right)$$

$$Y_n(\theta) = [I_{T_n}(\theta)]^{\frac{1}{2}} \left(\frac{\partial \log \varphi_{T_n}}{\partial \theta} \right)$$

$$U_n(\theta) = [I_n(\theta)]^{-\frac{1}{2}} (T_n - \theta) .$$

Enfin, nous faisons l'*hypothèse* 4.5. . La loi asymptotique de $(Z_n(\theta), U_n(\theta))$ existe et possède les moments d'ordre 1 et 2 finis.

On vérifie facilement les lemmes suivants

LEMME 4.1.1. –

$$(4.1.5.) \quad \text{(i)} \quad E_\theta(Z_n \mid T_n) = \left[\frac{I_{T_n}(\theta)}{I_n(\theta)} \right]^{1/2} Y_n(\theta) \quad \lambda^{(n)} \text{ p.p.}$$

$$(4.1.6.) \quad \text{(ii)} \quad I_{T_n}(\theta) \leqslant I_n(\theta) \ \forall \text{ n } \textit{l'égalité se produisant si et seulement si } T_n \textit{ est une statistique exhaustive}$$

$$\text{(iii)} \qquad E_\theta(Y_n(\theta)\, Z_n(\theta)) = \left[\frac{I_{T_n}(\theta)}{I_n(\theta)} \right]^{1/2}$$

LEMME 4.1.2. – *Soit* $U_n^c = U_n$ *si* $|U_n| \leqslant c$
$= 0$ *autrement.*

où c *est un point de continuité de la fonction de répartition limite de* U_n .

Soit $E_\theta [U_n^c Z_n] = \alpha_n^c$

$E_\theta [(U_n^c)^2] = V_n^c$

Alors,

(i) *Pour* c *fixé,* $\lim_{n \to \infty} \alpha_n^c = \alpha^c$ où α^c *est le moment produit relatif à la fonction de répartition limite de* (Z_n, U_n^c) .

(ii) *Pour* $\lim_{n \to \infty} V_n^c = V^c$ *où* V^c *est le deuxième moment relatif à la fonction de répartition limite de* U_n^c .

(iii) $\lim_{c \to \infty} \alpha^c = \alpha$, *le moment produit relatif à la fonction de répartition limite de* Z_n *et* U_n .

(iv) $\lim_{c \to \infty} V^c = V$ *le deuxième moment de la fonction de répartition limite de* U_n .

Les démonstrations de ces lemmes se font essentiellement comme dans Rao [46] Sec. 3.

PROPOSITION 4.1.1. — *Supposons que* T_n *est un estimateur sans biais de* θ *(en supposant que* θ *est la vraie valeur du paramètre) pour* $\forall\, n \in \mathbf{N}^*$

(4.1.8.) $$\liminf_{n \to \infty} \frac{I_{T_n}(\theta)}{I_n(\theta)} \geqslant \rho^2$$ *où* ρ *est la corrélation asymptotique de* Z_n *et* U_n .

DÉMONSTRATION. — On a

$$\mathrm{Var}_\theta (U_n^c)\, \mathrm{Var}_\theta \left\{ \left[\frac{I_{T_n}(\theta)}{I_n(\theta)} \right]^{1/2} Y_n(\theta) \right\} \geqslant$$

$$\geqslant \left\{ E_\theta \left\{ \left[\frac{I_{T_n}(\theta)}{I_n(\theta)} \right]^{1/2} U_n^c(\theta)\, Y_n(\theta) \right\} \right\}^2 =$$

$$= \left[E_\theta \left(Z_n(\theta)\, U_n^c(\theta) \right) \right]^2 = (\alpha_n^c)^2 .$$

Donc

$$\frac{(\alpha_n^c)^2}{V_n^c} = (\rho_n^c)^2 \leqslant \frac{I_{T_n}(\theta)}{I_n(\theta)}$$

puisque $\mathrm{Var}_\theta(Y_n(\theta)) = \mathrm{Var}_\theta(Z_n(\theta)) = 1$.

Donc $$\liminf_{n \to \infty} \frac{I_{T_n}(\theta)}{I_n(\theta)} \geqslant (\rho^c)^2$$

comme c est arbitraire lorsque $c \to \infty$ le résultat est obtenu.

PROPOSITION 4.1.2. — *Si* $|Z_n - \alpha_n - \beta_n U_n| \to 0$ *en probabilité quand* $n \to \infty$, *où* α_n *et* β_n *ne dépendent pas de l'échantillon et* $\beta_n \neq 0$, *alors*

$$\lim_{n \to \infty} \frac{I_{T_n}(\theta)}{I_n(\theta)} = 1 \quad . \tag{4.1.9.}$$

DÉMONSTRATION. — D'après le lemme 4.1.1. on a

$I_{T_n}(\theta) \leqslant I_n(\theta)$ et, par conséquent ,

$$\limsup_{n \to \infty} \frac{I_{T_n}(\theta)}{I_n(\theta)} \geqslant 1 \quad .$$

D'après l'hypothèse formulée dans la proposition, il résulte que la corrélation limite de Z_n et U_n est l'unité.

Donc, par la proposition 4.1.1.

$$\liminf_{n \to \infty} \frac{I_{T_n}(\theta)}{I_n(\theta)} \geqslant 1$$

D'où le résultat.

PROPOSITION 4.1.3. — *Pour tester l'hypothèse* $H_o : \theta = \theta_o$,

(i) $Z_n(\theta_o) \geq e_1$, (ii) $Z_n(\theta_o) \leq e_2$ *donnent respectivement les tests localement les plus puissants contre les alternatives* (ii) $\theta > \theta_o$ *et* (ii) $\theta < \theta_o$.

(iii) $|Z_n(\theta_o)| > e_3$ *donne un test localement le plus puissant sans biais contre* $\theta \neq \theta_o$. *Dans chaque cas, "e" est choisi de façon à satisfaire à la condition de seuil.*

Ces résultats sont des conséquences immédiates du lemme de Neyman-Pearson (voir ég. [43] ou [52] pour le cas des variables indépendantes).

4.2. Définition de l'efficacité asymptotique

La nécessité d'une reformulation du concept d'efficacité expliquée au paragraphe 4.0. nous conduit à définir l'efficacité asymptotique comme suit

DÉFINITION 4.2.1. – Un estimateur T_n de θ est dit asymptotiquement efficace pour θ si l'une des conditions suivantes est satisfaite

1/
(4.2.1.) $$\frac{I_{T_n}(\theta)}{I_n(\theta)} \to 1 \quad \text{lorsque } n \to \infty ,$$

2/
(4.2.2.) La corrélation asymptotique de Z_n et U_n est l'unité.

3/ Il existe deux constante α et β $(\neq 0)$ ne dépendant pas de l'échantillon, telles que

(4.2.3.) $$|Z_n - \alpha - \beta U_n| \to 0 \text{ en probabilité lorsque } n \to \infty .$$

Il faut bien préciser que les trois formes de la définition ne sont pas équivalentes. La forme 3/ est la plus restrictive mais a l'avantage d'être vérifiée plus facilement, même sous les hypothèses moins strictes. D'ailleurs, puisque la fonction de répartition limite de Z_n et U_n existe,

d'après l'hypothèse 4.5. avec les moments d'ordre 1 et 2 finis, il s'ensuit que $3/ \Longrightarrow 2/$. Il semble que la forme 2/ soit plus logique et plus appropriée pour le cas présent, mais il est assez difficile de le vérifier. De la proposition 4.2.2., de 3/ et de 2/ on déduit 1/ . La forme 1/ maintient l'idée originale, due à Fisher, de comparer les informations à la limite dans l'échantillon et dans la statistique.

Donc, dans ce qui va suivre, nous allons employer la définition sous la forme 3/ .

Voici quelques remarques concernant la définition ci-dessus

Remarque 1. — L'introduction de la forme 3/ de la définition ne nécessite pas l'hypothèse 4.5.

Remarque 2. — On n'a fait aucune hypothèse sur la fonction de coût.

Remarque 3. — Cette définition donne une formulation précise de ce que Wald [55] a demandé dans sa définition de l' "efficacité au sens large".

Remarque 4. — Dans le cas de variables aléatoires indépendantes, la discussion qui suit montre qu'une variante plus forte que la forme 3/ de cette définition implique les exigences de Bahadur [8] dans sa recherche d'une nouvelle définition de l'efficacité asymptotique par l'intermédiaire de "variances effectives" (Effective variances").

La définition donnée par Bahadur concerne l'efficacité relative entre deux estimateurs de θ en faisant la comparaison de leur concentration dans les intervalles fixes autour de la vraie valeur de θ . Si l'on compare cette concentration à une quantité indépendante de la méthode d'estimation, celle-ci donne aussi une "mesure de la valeur" d'un estimateur et conserve les propriétés correspondant à la définition de Bahadur. D'abord, l'écart-type effectif (effective standard deviation) est défini comme suit ([8] , p. 230).

DÉFINITION. — Soit $g(\theta)$, une fonction réelle de θ . Pour chaque statistique $T_n(x_1,\ldots,x_n)$ à valeurs réelles que nous notons, pour simplifier T_n, pour chaque θ et chaque $\varepsilon > 0$, supposons que $\tau = \tau_g(T_n ; \varepsilon ; \theta)$ est la solution de l'équation

$$\frac{1}{\sqrt{2\pi}} \int_{\{x : |x| > \frac{\varepsilon}{\tau}\}} e^{-1/2\, x^2}\, dX = P\{|T_n - g(\theta)| \geqslant \varepsilon ; \theta\} ,$$

où $0 \leqslant \tau \leqslant \infty$. Alors τ est appelé écart-type effectif de T_n, quand

a) T_n est un estimateur de $g(\theta)$,

b) θ est la vraie valeur du paramètre et

c) quand on a besoin de calculer $P(|T_n - g(\theta)| \geqslant \varepsilon ; \theta)$.

Nous allons donc donner une définition comparant la variance effective d'un estimateur T_n de θ avec $\tau^2(n, \varepsilon ; \theta)$ celle de la fonction de vraisemblance qui est égale à $[n I(\theta)]^{-1}$, où $I(\theta)$ est l'information dans le cas d'une observation.

DÉFINITION 4.2.2. – Un estimateur T_n est dit asymptotiquement efficace pour θ, si

$$(4.2.4.) \qquad \limsup_{\varepsilon \to 0} \limsup_{n \to \infty} \left\{ \frac{\tau^2(n, \varepsilon ; \theta)}{\tau^2(T_n, \varepsilon ; \theta)} \right\} = 1$$

c'est-à-dire

$$\limsup_{\varepsilon \to 0} \limsup_{n \to \infty} \left\{ \frac{-2 \log P(|T_n - \theta| > \varepsilon ; \theta)}{\varepsilon^2\, n\, I(\theta)} \right\} = 1 .$$

Le résultat suivant est connu (Rao [48]). Si un estimateur T_n de θ tel que, pour chaque $\varepsilon > 0$

$$P\left\{ |T_n - \theta_o - \frac{S_n(\theta_o)}{nI(\theta_o)}| \geqslant \frac{\varepsilon\, S_n(\theta_o)}{n\, I(\theta_o)} ; \theta_o \right\} < \eta^n$$

$$\text{ou} \quad \left\{ P \left| T_n - \theta_o - \frac{S_n(\theta_o)}{n\, I(\theta_o)} \right| > \varepsilon\, |T_n - \theta_o| \, \theta_o \right\} < \eta^n ,$$

où $S_n(\theta_o) = \left(\frac{\partial \log f_n}{\partial \theta}\right)_{\theta = \theta_o}$ et $0 < \eta < 1$ qui peut dépendre de θ et ε ; si en outre $P(T_n \geqslant \theta\ ;\ \theta)$ pour $\theta \leqslant \theta_o$ et $P(T_n \leqslant \theta\ ;\ \theta)$ pour $\theta \geqslant \theta_o$ ont des bornes inférieures à l'unité, alors

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \left\{ \frac{-2 \log P(|T_n - \theta| \geqslant \varepsilon\ ;\ \theta)}{n \varepsilon^2 I(\theta)} \right\} = 1 .$$

Un estimateur qui satisfait aux conditions ci-dessus est appelé "estimateur asymptotiquement fortement efficace de premier ordre (Strong first order efficient)". On peut noter que ces conditions sont du même genre que celles qui figurent dans la définition 4.2.1. forme 3/ et si un estimateur satisfait à ces conditions, il est automatiquement efficace d'après la définition 4.2.1.

PROPOSITION 4.2.1. – *Si un estimateur* T_n *est asymptotiquement fortement efficace du premier ordre pour la vraie valeur du paramètre* θ , *alors* T_n *est efficace au sens de la définition* 4.2.2. *et donc celle de Bahadur.*

DÉMONSTRATION. – Soit V_n un estimateur convergent en probabilité de θ . L'efficacité asymptotique de T_n par rapport à V_n est définie par Bahadur comme étant

$$(4.2.7.) \quad \limsup_{\varepsilon \to 0} \limsup_{n \to \infty} \left\{ \frac{\tau^2(T_n, \varepsilon\ ;\ \theta)}{\tau^2(V_n, \varepsilon\ ;\ \theta)} \right\} =$$

$$= \limsup_{\varepsilon \to 0} \limsup_{n \to \infty} \left\{ \frac{\tau^2(T_n, \varepsilon\ ;\ \theta)}{\tau^2(n, \varepsilon\ ;\ \theta)} \quad \frac{\tau^2(n, \varepsilon\ ;\ \theta)}{\tau^2(V_n, \varepsilon\ ;\ \theta)} \right\} =$$

$$= \limsup_{\varepsilon \to 0} \limsup_{n \to \infty} \left\{ \frac{\tau^2(n, \varepsilon\ ;\ \theta)}{\tau^2(V_n, \varepsilon\ ;\ \theta)} \right\} \leqslant 1$$

puisqu'un estimateur fortement asymptotiquement efficace du premier ordre est efficace d'après la définition 4.2.2., et qu'il résulte que

$$\frac{\tau^2(T_n, \varepsilon\ ;\ \theta)}{\tau^2(n, \varepsilon\ ;\ \theta)} \to 1 \quad \text{lorsque } n \to \infty\ , \quad \varepsilon \to 0$$

et $\liminf_{\varepsilon \to 0} \liminf_{n \to \infty} \frac{\tau^2 (V_n, \varepsilon ; \theta)}{\tau^2 (n, \varepsilon ; \theta)} \leqslant 1$ (Bahadur)

Ceci démontre que T est efficace au sens de Bahadur.

Maintenant faisons quelques hypothèses supplémentaires, dans le cas général

Hypothèse 4.6. – Pour chaque θ , $\left[\mathrm{Var}_\theta \left\{\frac{\partial^2 \log f_n}{\partial \theta^2}\right\}\right]^{1/2}$ divisée par $E_\theta \left(\frac{\partial^2 \log f_n}{\partial \theta^2}\right)$ tend vers 0 lorsque $n \to \infty$.

Hypothèse 4.7. – $\frac{\partial^3 f_n}{\partial \theta^3}$ existe et il existe un $\delta > 0$ tel que

$$\frac{1}{I_n(\theta)} \; E_\theta \left\{ \sup_{|\theta - \theta'| < \delta} \left| \left(\frac{\partial^3 \log f_n}{\partial \theta^3}\right)_{\theta = \theta'} \right| \right\} \leqslant \gamma(\theta) < \infty$$

où $\gamma(\theta)$ est une fonction de θ .

PROPOSITION 4.2.2. – *Sous les hypothèses* 4.1. *à* 4,4., 4,6,, 4.7.

(i) *avec une probabilité qui tend vers* 1 *quand* $n \to \infty$ *l'équation de vraisemblance a une racine* $\hat{\theta}_n$

(ii) $\hat{\theta}_n$ *est un estimateur convergent en probabilité de la vraie valeur de paramètre* θ .

(iii) $\hat{\theta}_n$ *est asymptotiquement efficace.*

La démonstration se fait essentiellement comme dans le cas classique (Cramer [19] ou Wald [55]) ou comme dans les cas particuliers ci-dessous.

4.3. Estimation pour des processus de Markov

Dans les cas particuliers, lorsqu'on considère des processus spécifiques, comme dans ce qui va suivre, il est souvent possible d'affaiblir les hypothèses et en même temps d'obtenir des résultats plus précis.

4.3.1. – Nous nous plaçons dans le cas d'un processus de Markov à temps discret avec une seule classe ergodique d'états sans sous-classe cyclique. Les probabilités de transition sont données par

$$P(\xi, A ; \theta) = \Pr\{X_{n+1} \in A \mid X_n = \xi ; \theta\}$$

$\forall A \in \mathcal{S}$ et $\theta \in \Theta$.

Soit $P(. ; \theta)$ la probabilité absolue stationnaire telle que

$$P(A, \theta) = \int_{\mathcal{X}} P(\xi, A ; \theta)\, P(d\xi ; \theta) \qquad \forall A \in \mathcal{S}$$

Nous faisons les hypothèses suivantes

Hypothèse D.1. – (Condition d'identifiabilité). Les probabilités de transition et la probabilité absolue stationnaire sont telles que

$$P(\xi, . ; \theta_1) = P(\xi, . ; \theta_2)$$

et $$P(. ; \theta_1) = P(. ; \theta_2)$$

si et seulement si $\theta_1 = \theta_2$.

Hypothèse D.2. – Il existe une mesure σ-finie, λ, sur $\mathcal{S}$ par rapport à laquelle la probabilité de transition et la probabilité absolue stationnaire admettent des densités de Radon-Nikodym $f(x, y ; \theta)$ et $f(x ; \theta)$ données par

$$P(x, A ; \theta) = \int_A f(x, y ; \theta)\, \lambda(dy)$$

$$P(A ; \theta) = \int_A f(x ; \theta)\, \lambda(dx) \qquad \forall A \in \mathcal{S} .$$

Hypothèse D.3. – $\forall\, x \in \mathcal{S}$ l'ensemble des y tels que $f(x, y ; \theta) > 0$ ne dépend pas de θ . Pour tout x et y λ-presque partout $\in \mathcal{X}$

$$\frac{\partial f(x, y ; \theta)}{\partial \theta} \quad , \quad \frac{\partial^2 f(x, y ; \theta)}{\partial \theta^2}$$

existent et sont continues en $\theta \in \Theta$. De plus,

$$\frac{\partial^i}{\partial \theta^i} \int_{\mathcal{X}} f(x, y ; \theta) \lambda(dy) = \int_{\mathcal{X}} \frac{\partial^i (f(x, y ; \theta) \lambda(dy)}{\partial \theta^i}$$

pour $i = 1, 2$.

Il s'ensuit que $E_\theta \left(\frac{\partial \log f(x, y ; \theta)}{\partial \theta} \right) = 0 , \quad \forall \theta \in \Theta$

et $E_\theta \left[\left(\frac{\partial \log f(x, y ; \theta)}{\partial \theta} \right)^2 \right] = -E\left(\frac{\partial^2 \log f(x, y ; \theta)}{\partial \theta^2} \right)$

ce que nous notons $I(\theta)$.

(Dans ce qui va suivre, les espérances mathématiques sont calculées en prenant $P(. ; \theta)$ comme probabilité initiale du processus comme dans le chapitre 1) .

On suppose que $I(\theta)$ est continue en $\theta \in \Theta$, et qu'il existe une constante $0 < k \leqslant I(\theta)$, $\forall \theta \in \Theta$.

Hypothèse D.4. – Pour chaque $\theta \in \Theta$, il existe un voisinage $V(\theta)$ tel que, pour chaque $\theta' \in V(\theta)$,

$$\left| \left(\frac{\partial^2 \log f(x,y ; \theta)}{\partial \theta^2} \right)_\theta - \left(\frac{\partial^2 \log f(x,y ; \theta)}{\partial \theta^2} \right)_{\theta'} \right| \leqslant |\theta - \theta'| \, G(x,y)$$

où $0 \leqslant G(x,y)$ est telle que $E_\theta(G(x,y)) < \infty , \forall \theta$.

Hypothèse D.5. – Pour chaque $\theta \in \Theta$, la probabilité absolue stationnaire est telle que

$$P(x, . ; \theta) \ll P(. ; \theta) \ \forall x \in \mathcal{X} .$$

(On peut noter qu'il suffit de supposer que les conditions de l'hypothèse D.3. sont remplies dans un voisinage de la vraie valeur de θ) .

Avec $\underset{\sim}{X}^{(n)} = (X_1, \ldots, X_n)$, la vraisemblance s'écrit

$$L_n(\underset{\sim}{X}^{(n)} ; \theta) = f(X_1 ; \theta) \prod_{i=1}^{n-1} f(X_i, X_{i+1} ; \theta)$$

où $\log L_n(\underset{\sim}{X}^{(n)} ; \theta) = \log f(X_1 ; \theta) + \sum_{i=1}^{n-1} \log f(X_i, X_{i+1} ; \theta)$.

Comme dans le chapitre 1, en convervant seulement le terme dominant, on écrit le logarithme de la vraisemblance

$$\log L_n(\underset{\sim}{X}^{(n)} ; \theta) \approx \sum_{i=1}^{n-1} \log f(X_i, X_{i+1} ; \theta)$$

Donc l'équation de vraisemblance s'écrit

$$\frac{\partial \log L_n}{\partial \theta} = 0 .$$

PROPOSITION 4.3.1. – *Sous les hypothèses* D.1 - D.5

(i) *Avec une probabilité qui tend vers* 1 *quand* $n \to \infty$. *L'équation de vraisemblance a une racine* $\hat{\theta}_n(\underset{\sim}{X}^{(n)})$.

(ii) $\hat{\theta}_n(\underset{\sim}{X}^{(n)})$ *est un estimateur convergent de la vraie valeur du paramètre* θ . *Si* θ *est la vraie valeur du paramètre* $\hat{\theta}_n(\underset{\sim}{X}^{(n)}) \to \theta$ *en probabilité quand* $n \to \infty$.

(iii) $\hat{\theta}_n(\underset{\sim}{X}^{(n)})$ *est asymptotiquement efficace.*

(iv) $\hat{\theta}_n(\underset{\sim}{X}^{(n)})$ *suit asymptotiquement la loi gaussienne d'espérance mathématique* θ *et de variance* $[(n-1)\ I(\theta)]^{-1}$.

Avant de donner la démonstration de cette proposition, nous rappelons tout d'abord un résultat de Billingsley ([13] , p.6) , qui simplifie les calculs

" Si l'hypothèse D.5. est satisfaite, et si la probabilité initiale est la probabilité absolue stationnaire, le processus est transitif métrique (metrically transitive). Quelle que soit la probabilité initiale, si $\varphi(X, Y)$ est une fonction mesurable du couple (X, Y) et $E_\theta(|\varphi(X, Y)|) < \infty$, alors

$$\lim_{n \to \infty} \frac{1}{n} \sum_{1}^{n} \varphi(X_1 , X_{i+1}) = E_\theta(|\varphi(X_1, X_2)|)$$

presque sûrement " .

DÉMONSTRATION DE LA PROPOSITION. – Ecrivons

$$\frac{\partial \log L_n}{\partial \theta} = \left(\frac{\partial \log L_n}{\partial \theta}\right)_{\theta = \theta_0} + (\theta - \theta_0)\left[\left(\frac{\partial^2 \log L_n}{\partial \theta^2}\right)_{\theta = \theta_0} + \left(\frac{\partial^2 \log L_n}{\partial \theta^2}\right)_{\theta = \theta_0 + \alpha(\theta - \theta_0)} - \left(\frac{\partial^2 \log L_n}{\partial \theta^2}\right)_{\theta = \theta_0}\right]$$

où $0 < |\alpha| < 1$.

Du théorème de Billingsley ci-dessus, on a $\frac{1}{(n-1)} \frac{\partial \log L_n}{\partial \theta} \to 0$

$$(4.3.10.) \quad \frac{1}{(n-1)} \frac{\partial^2 \log L_n}{\partial \theta^2} \longrightarrow -I(\theta)$$

$$\frac{1}{(n-1)} \sum_{i=1}^{n-1} G(X_i, X_{i+1}) \to (\mathbf{E}_\theta (G(X_1, X_2))) = \mathbf{M} < \infty ,$$

presque sûrement quand $n \to \infty$.

L'équation de vraisemblance s'écrit

$$\frac{1}{(n-1)} \frac{\partial \log L_n}{\partial \theta} = \frac{1}{(n-1)} \left(\frac{\partial \log L_n}{\partial \theta}\right)_{\theta = \theta_0} + \frac{1}{(n-1)} (\theta - \theta_0)\left(\frac{\partial^2 \log L_n}{\partial \theta^2}\right)_{\theta = \theta_0}$$

$$(4.3.11) \qquad + \frac{1}{(n-1)} \alpha_n . (\theta - \theta_0)^2 \sum_{i=1}^{n-1} G(X_i, X_{i+1})$$

où $|\alpha_n| < 1$, $\forall$ n .

Soit $\varepsilon > 0$ un nombre quelconque. Choisissons $\delta = \delta(\varepsilon) > 0$ tel que le voisinage $|\theta - \theta_0| < \delta$ soit contenu dans le voisinage $V(\theta_0)$ de θ_0 pour lequel l'hypothèse D.4. est vérifiée.

Soient S_i $(i = 1, 2, 3)$ les ensembles de points $\underset{\sim}{X}^{(n)}$ tels que

$$S_1 : \left\{ \underset{\sim}{X}^{(n)} \;\middle|\; \frac{1}{(n-1)} \left(\frac{\partial \log L_n}{\partial \theta}\right)_{\theta = \theta_0} > \delta^2 \right\}$$

$$S_2 : \left\{ \underset{\sim}{X}^{(n)} \;\middle|\; \left| \frac{1}{(n-1)} \left(\frac{\partial^2 \log L_n}{\partial \theta^2} \right) \theta = \theta_o + I(\theta_o) \right| > \delta \right\}$$

$$S_3 : \left\{ \underset{\sim}{X}^{(n)} \;\middle|\; \frac{1}{(n-1)} \sum_{i=1}^{n-1} G(X_i, X_{i+1}) > M + 1 \right\}$$

et

(4.3.12) $S = \complement (S_1 \cup S_2 \cup S_3)$. Alors pour $n > n_o(\delta, \varepsilon)$ $Pr(S ; \theta_o) > 1 - \varepsilon$.

Par conséquent, pour $\theta = \theta \pm \delta$, la somme des premier et troisième termes dans l'expression (4.3.11.) est inférieure à $(M + 2)\delta^2$ avec la probabilité $> 1 - \varepsilon$ dès que $n \geqslant n_o(\delta, \varepsilon)$. Maintenant, si $\delta(\varepsilon)$ est aussi choisi tel que $\delta(\varepsilon) < \dfrac{I(\theta_o)}{(M+3)}$, alors le signe de cette expression ne dépend que du deuxième terme c'est-à-dire

$$(4.3.13) \qquad \frac{\partial \log L_n}{\partial \theta} \quad \begin{array}{ll} > 0 & \text{pour } \theta = \theta_o - \delta \\ < 0 & \text{pour } \theta = \theta_o + \delta \end{array} .$$

Puisque $\dfrac{\partial \log L_n}{\partial \theta}$ est dérivable en θ pour tout $\underset{\sim}{X}^{(n)}$, $\lambda^{(n)}$ -presque partout, il existe une racine $\hat{\theta}_n(\underset{\sim}{X}^{(n)})$ pour l'équation de vraisemblance

$$\frac{1}{(n-1)} \frac{\partial \log L_n}{\partial \theta} = 0 \text{ dans le voisinage } (\theta_o - \delta, \theta_o + \delta)$$

pour $n \geqslant n_o(\delta, \varepsilon)$ avec une probabilité $> 1 - \varepsilon$. Comme δ et ε sont arbitraires, ceci démontre (i) et (ii).

En remplaçant les $\hat{\theta}_n(\underset{\sim}{X}^{(n)}$ et en divisant par $I(\theta_o)$, on a

$$\frac{1}{(n-1)I(\theta_o)} \left(\frac{\partial \log L_n}{\partial \theta} \right)_{\theta = \theta_o} + (\hat{\theta}_n - \theta_o) \frac{1}{(n-1)I(\theta_o)} \left(\frac{\partial^2 \log L_n}{\partial \theta^2} \right) \theta = \theta_o +$$

$$(4.3.14) \qquad + \frac{1}{(n-1)I(\theta_o)} (\hat{\theta}_n - \theta_o) U_n = 0$$

où $U_n = \left[\left(\frac{\partial^2 \log L_n}{\partial \theta^2} \right)_{\theta = \theta_o + \alpha_n (\hat{\theta}_n - \theta_o)} - \left(\frac{\partial^2 \log L_n}{\partial \theta^2} \right)_{\theta = \theta_o} \right]$,

tend vers 0 en probabilité quand $n \to \infty$.

c'est-à-dire

$$\frac{1}{\sqrt{(n-1)\, I(\theta_o)}} \left(\frac{\partial \log L_n}{\partial \theta} \right)_{\theta = \theta_o} - \sqrt{(n-1)\, I(\theta_o)}\, (\hat{\theta}_n - \theta_o) + \xi_n = 0$$

où $\xi_n \to 0$ en probabilité quand $n \to \infty$ puisque

$\frac{1}{(n-1)\, I(\theta_o)} \left(\frac{\partial^2 \log L_n}{\partial \theta^2} \right) \longrightarrow -1$ et $(\hat{\theta}_n - \theta_o)\, U_n \longrightarrow 0$ en probabilité.

Ceci démontre (iii) .

Enfin, puisque $\frac{1}{\sqrt{(n-1)\, I(\theta_o)}} \left(\frac{\partial \log L_n}{\partial \theta} \right)_{\theta = \theta_o}$ suit asymptotiquement la loi normale d'espérance mathématique 0 et de variance unité, on a la même propriété pour $\sqrt{(n-1)\, I(\theta_o)}\, (\hat{\theta}_n - \theta_o)$. Ceci complète la démonstration de la proposition. En passant, signalons que les hypothèses peuvent être affaiblies.

4.3.2. – Maintenant, supposons que $\mathcal{X}$ ne comporte qu'un nombre fini d'états notés 1,...,m . Dans ce cas, à la place de la densité de probabilité de transition, nous avons une matrice $P(\theta) = (p_{ij}(\theta))$ d'ordre $m \times m$ avec les probabilités de transition données par

$$p_{ij}(\theta) = \Pr \{ X_{n+1} = j \mid X_n = i \,;\, \theta \}, \quad \theta \in \Theta \tag{4.3.15}$$

pour $i, j = 1,...,m$.

Supposons qu'il n'existe qu'une seule classe ergodique. Alors il s'ensuit qu'il existe un ensemble de probabilités stationnaires $\Pi_i(\theta) > 0$, $i = 1,...,m$, données par

$$\lim_{n \to \infty} p_{ij}^{(n)}(\theta) = \Pi_j(\theta) > 0$$

indépendant de i pour tout $i, j = 1,\ldots,m$ où $p_{ij}^{(n)}(\theta)$ est la probabilité que l'évolution qui commence en i passe par l'état j après n étapes. Il est bien connu que $(p_{ij}^{(n)}(\theta)) = P^n(\theta)$, $\forall\, \theta \in \Theta$, et $\forall\, n \in \mathbf{N}^*$.

Le processus de Markov décrit ici est parfois appelé "une chaîne de Markov régulière (Fréchet [22])" . On sait que la loi asymptotique stationnaire d'une telle chaîne de Markov est multinomiale. Donc, les analogies entre les méthodes sont prises en compte dans les démonstrations.

Nous faisons les hypothèses suivantes

Hypothèse F. – (condition d'identifiabilité)

$$p_{ij}(\theta_1) \neq p_{ij}(\theta_2) \quad \text{pour} \quad \theta_1 \neq \theta_2$$

au moins pour un couple $(i,j) \in (\mathcal{X} \times \mathcal{X})$ et

$$\Pi_i(\theta_1) \neq \Pi_i(\theta_2) \quad \text{au moins pour un} \quad i \in \mathcal{X} .$$

Hypothèse F.2. – $p_{ij}(\theta)$ sont dérivables et les dérivées sont continues pour la vraie valeur du paramètre (disons θ_o) .

Hypothèse F.3. – L'information de Fisher

$$I(\theta) = \sum_{i,j=1}^{m} \Pi_i(\theta)\, p_{ij}(\theta) \left(\frac{1}{p_{ij}(\theta)} \quad \frac{dp_{ij}(\theta)}{d\theta} \right)^2$$

est positive pour tout θ .

Supposons qu'un échantillon $X_1,\ldots,X_n$ est tiré du processus. Soit n_{ij} le nombre de fois que $X_r = i$, $X_{r+1} = j$ pour $r = 1, \ldots, n\text{-}1$ et $i,j = 1, \ldots, m$. La vraisemblance de l'échantillon est donnée par

$$(4.3.16) \qquad L_n(\underset{\sim}{X}^{(n)} ; \theta) = P(X_1 ; \theta) \prod_{i,j=1}^{m} p_{ij}^{n_{ij}}(\theta)$$

ou $$\log L_n(\underset{\sim}{X}^{(n)} ; \theta) = \log P(X_1 ; -\theta) + \sum_{i,j=1}^{m} n_{ij} \log p_{ij}(\theta)$$

comme dans le cas précédent, on peut écrire

$$(4.3.17) \qquad \log L_n(\underset{\sim}{X}^{(n)} ; \theta) \approx \sum_{i,j=1}^{m} n_{ij} \log p_{ij}(\theta) .$$

L'équation de vraisemblance est

$$(4.3.18) \qquad \frac{\partial \log L_n(\underset{\sim}{X}^{(n)} ; \theta)}{\partial \theta} = \sum_{i,j=1}^{m} n_{ij} \frac{1}{p_{ij}(\theta)} \frac{dp_{ij}(\theta)}{d\theta} = 0$$

PROPOSITION 4.3.2. – *Sous les hypothèses* F.1 - F.3 *l'équation de vraisemblance possède avec une probabilité qui tend vers* 1 *lorsque* $n \to \infty$, *une racine* $\hat{\theta}_n(\underset{\sim}{X}^{(n)})$ *qui est convergente en probabilité et asymptotiquement efficace pour* θ_o . $\hat{\theta}_n(\underset{\sim}{X}^{(n)})$ *suit asymptotiquement la loi gaussienne d'espérance mathématique* θ_o *et de variance* $[n I(\theta_o)]^{-1}$.

On peut remarquer que les hypothèses faites ici sont plus faibles que dans le cas précédent.

DÉMONSTRATION. – Tout d'abord remarquons que la loi exacte de (n_{ij}) est établie par Whittle [56] , qui a aussi montré que

$$\bar{p}_{ij} = \frac{n_{ij}}{n} \longrightarrow \pi_i(\theta)\, p_{ij}(\theta) \quad \text{presque sûrement}$$

(4.3.19) et

$$\frac{n_{ij}}{n_i} \longrightarrow p_{ij}(\theta) \quad \text{presque sûrement}$$

lorsque $n \to \infty$ où $n_i = \sum_{j=1}^{m} n_{ij}$, $i = 1, \ldots, m$.

La démonstration de la première partie se fait comme dans le cas multinomial et nous allons seulement signaler les points essentiels.

De l'inégalité de Jensen, on a

$$(4.3.20) \qquad \sum_{i,j=1}^{m} \pi_i(\theta_o)\, p_{ij}(\theta_o) \log p_{ij}(\theta_o) > \sum_{i,j=1}^{m} \pi_i(\theta_o)\, p_{ij}(\theta_o) \log p_{ij}(\theta),$$

l'inégalité stricte se produisant à cause de l'hypothèse F.1. pour $|\theta - \theta_o| \leq \delta$ pour un nombre $\delta > 0$.

Puisque $\overline{p}_{ij} = \frac{n_{ij}}{n} \longrightarrow \pi_i(\theta_o)\, p_{ij}(\theta_o)$ p.s. quand $n \to \infty$, nous avons, pour n assez grand

$$(4.3.21) \qquad \sum_{i,j=1}^{m} \overline{p}_{ij} \log p_{ij}(\theta_o) \geqslant \sum_{i,j=1}^{m} \overline{p}_{ij} \log p_{ij}(\theta)$$

pour $|\theta - \theta_o| = \delta$. Donc, de l'hypothèse F.2., il découle qu'il existe un point $\hat{\theta}_n$ dans l'intervalle $|\theta - \theta_o| \leqslant \delta$ pour lequel

$$\sum_{i,j=1}^{m} \overline{p}_{ij} \log p_{ij}(\theta) \text{ est un maximum local.}$$

Comme les $p_{ij}(\theta)$ sont dérivables, la dérivée de $\sum_{i,j=1}^{m} \overline{p}_{ij} \log p_{ij}(\theta)$ s'annule pour $\theta = \hat{\theta}_n$.

Comme δ est arbitraire, $\hat{\theta}_n$ est un estimateur convergent de θ_o .

L'équation de vraisemblance s'écrit

$$(4.3.22) \qquad n \sum_{i,j=1}^{m} \frac{(\overline{p}_{ij} - \pi_i(\theta_o)\, p_{ij}(\theta_o))}{p_{ij}(\hat{\theta}_n)} \left(\frac{dp_{ij}(\theta)}{d\theta} \right)_{\theta = \hat{\theta}_n}$$

$$- n \sum_{i,j=1}^{m} \frac{\pi_i(\theta_o)\,(p_{ij}(\hat{\theta}_n) - p_{ij}(\theta_o))}{p_{ij}(\hat{\theta}_n)} \left(\frac{dp_{ij}(\theta)}{d\theta} \right)_{\theta = \hat{\theta}_n} = 0$$

Mais

$$(4.3.23) \qquad p_{ij}(\hat{\theta}_n) = p_{ij}(\theta_o) + (\hat{\theta}_n - \theta_o) \left(\frac{dp_{ij}(\theta)}{d\theta} \right)_{\theta = \theta'}$$

où $|\theta' - \theta_o| < |\hat{\theta}_n - \theta_o|$.

Donc l'équation (4.3.22) peut s'écrire

$$(4.3.24) \qquad n \sum_{i,j=1}^{m} \frac{(\bar{p}_{ij} - \pi_i(\theta_o) p_{ij}(\theta_o))}{p_{ij}(\hat{\theta}_n)} \left(\frac{dp_{ij}(\theta)}{d\theta}\right)_{\theta = \hat{\theta}_n} -$$

$$- n(\hat{\theta}_n - \theta_o) \sum_{i,j=1}^{m} \frac{\pi_i(\theta_o)}{p_{ij}(\hat{\theta}_n)} \left(\frac{dp_{ij}(\theta)}{d\theta}\right)_{\theta=\theta'} \left(\frac{dp_{ij}(\theta)}{d\theta}\right)_{\theta=\hat{\theta}_n} = 0$$

Pour n assez grand, puisque $\hat{\theta}_n$ est un estimateur convergent de θ_o, on a

$$\sum \frac{\pi_i(\theta_o)}{p_{ij}(\hat{\theta}_n)} \left(\frac{dp_{ij}(\theta)}{d\theta}\right)_{\theta=\theta'} \left(\frac{dp_{ij}(\theta)}{d\theta}\right)_{\theta=\hat{\theta}_n} \approx -I(\theta_o)$$

et

$$n \sum_{i,j=1}^{m} \frac{(\bar{p}_{ij} - \pi_i(\theta_o) p_{ij}(\theta_o))}{p_{ij}(\hat{\theta}_n)} \left(\frac{dp_{ij}(\theta)}{d\theta}\right) \approx Z_n(\theta_o)$$

et par conséquent, on a

$$\left| \frac{Z_n(\theta_o)}{\sqrt{nI(\theta_o)}} - \sqrt{nI(\theta_o)}\,(\hat{\theta}_n - \theta_o) \right| \to 0$$

en probabilité quand $n \to \infty$, ce qui démontre que $\hat{\theta}_n$ est asymptotiquement efficace. En outre, comme dans la proposition précédente, puisque le premier terme suit une loi gaussienne d'espérance mathématique 0 et de variance unité, il en est de même du deuxième, c'est-à-dire $\hat{\theta}_n$ suit asymptotiquement la loi normale d'espérance mathématique θ_n et de variance $[n I(\theta_o)]^{-1}$.

Ceci termine la démonstration.

Remarque. — Comme dans le cas multinomial, si l'on fait l'hypothèse plus forte que

$$\inf_{|\theta - \theta_o| > \delta} \sum_{j=1}^{m} p_{ij}(\theta_o) \log \frac{p_{ij}(\theta_o)}{p_{ij}(\theta)} \gg \varepsilon > 0$$

pour tout $i = 1,\ldots,m$, "condition dite d'identifiabilité forte", alors il s'ensuit que la racine $\hat{\theta}_n$ est un maximum absolu.

4.3.3. – Maintenant nous allons considérer le cas du processus de Markov multiple. On voit que le cas des processus de Markov multiples peut se ramener aux cas traités ci-dessus.

Prenons ici le cas où $\mathcal{X}$ a un nombre fini d'états. Une chaîne de Markov multiple peut être réduite comme une chaîne de Markov simple, si l'on emploie la technique d'introduction des vecteurs aléatoires (Doob. ch. IV). Alors les calculs faits dans le paragraphe 4.3.2. s'appliquent.

Soit donc un processus de Markov $\{X_n, n \in \mathbf{N}\}$ stationnaire d'ordre r avec les probabilités de transition stationnaires.

(4.3.25)

$$p_{i_{n-r},i_{n-r-1},\ldots,i_{n-1}\,;\,i_n} = \Pr\{X_n = i_n \mid X_{n-1} = i_{n-1} \ldots X_{n-r} = i_{n-r}\}$$

où $i_{n-r},\ldots,i_n \in \mathcal{X}$.

c'est-à-dire la probabilité conditionnelle de X_n ne dépend que des r états précédents $X_{n-1},\ldots,X_{n-r}$.

Soit P la matrice des probabilités stationnaires

$$p_{i_{n-r},\ldots,i_{n-1}\,;\,i_n} \qquad \text{où} \quad i_{n-r},\ldots,i_n \in \mathcal{X} .$$

Considérons l'espace produit $\mathcal{X}^{(r)}$. Soit $\underset{\sim}{Y}_n$ un vecteur aléatoire $(X_n,\ldots,X_{n+r-1})$ à r composantes. Alors $\{\underset{\sim}{Y}_n\}_{n \in \mathbf{N}^*}$ est une chaîne de Markov simple avec les probabilités de transition stationnaires

(4.3.26) $$q_{w,v} = \Pr\left\{\underset{\sim}{Y}_{n+1} = (v_1,\ldots,v_r) \;\middle|\; \underset{\sim}{Y}_n = (w_1,\ldots,w_r)\right\} =$$

$$= \begin{cases} p_{w_1,\ldots,w_r\,;\,v_r} & \text{si } w_{i+1} = v_i \text{ pour } i = 1,\ldots,r-1 ; \\ 0 & \text{sinon.} \end{cases}$$

à valeurs dans des états $\mathcal{X}^{(r)}$ qui comporte m^r r-uples.

Notons la matrice de probabilité de transition stationnaire $Q = (q_{w,v})$ qui sera une matrice d'ordre $m^r \times m^r$ avec les éléments données ci-dessus.

Un échantillon de taille $(n+r)$ du processus original est équivalent à un échantillon de taille n du processus associé.

Maintenant, supposons que les probabilités p_{ij} dépendent d'un paramètre $\theta \in \Theta$ comme ci-dessus et que l'on veuille estimer θ à partir d'un échantillon du processus.

Considérons un échantillon $\underset{\sim}{X}^{(n+r)} = (X_1,\ldots,X_{n+r})$ de taille $(n+r)$ du processus. La vraisemblance de $\underset{\sim}{X}^{(n+r)}$ s'écrit

(4.3.27)

$$L_n(.;\theta) = L(\underset{\sim}{X}^{(n+r)};\theta) = P(X_1;\theta)\, P(X_2 \mid X_1;\theta) \ldots$$

$$\ldots P(X_r \mid X_{r-1} \ldots X_1;\theta) \prod_{i=1}^{n} P(X_{r+i} \mid X_i,\ldots,X_{r+i-1};\theta) =$$

$$= P(X_1;\theta) \ldots P(X_r \mid X_{r-1},\ldots,X_1;\theta) \prod_{v,w} q_{w,v}^{n_{w,v}}(\theta)$$

où évidemment le produit s'étend à tous les couples $w,v \in \mathcal{X}^{(r)}$ tels que

si $\quad v = (v_1,\ldots,v_r) \quad$ et $\quad w = (w_1,\ldots,w_r)$

Alors $w_{i+1} = v_i$ pour $i = 1,\ldots,r-1$ et $n_{w,v}$ est la fréquence de passage en une étape de w à v dans cet échantillon.

Comme dansle paragraphe 4.3.2., toujours pour les mêmes raisons, en gardant le terme dominant.

$$\log L_n(.;\theta) \simeq \sum_{w,v} n_{w,v} \log q_{w,v}(\theta) \tag{4.3.28}$$

qui est de la même forme que celle du (4.3.17) pour le cas d'un processus de Markov simple. Donc, les résultats obtenus dans le cas simple s'appliquent ici.

Nous allons résumer cette discussion par la

PROPOSITION 4.3.3. — *Sous les hypothèses* F.1.- F.3., *l'équation de vraisemblance possède, avec une probabilité qui tend vers* 1 *quand* $n \to \infty$, *une racine* $\hat{\theta}_n$ *qui est un estimateur convergent de la vraie valeur du paramètre* θ *et qui est asymptotiquement efficace. De plus* $\hat{\theta}_n$ *suit asymptotiquement la loi normale d'espérance mathématique* θ_o *et de variance* $[nI(\theta_o)]^{-1}$.

4.3.4. Processus de Markov à temps continu

Dans ce paragraphe, avec des processus de Markov à temps continu et de type complètement discontinu, on voit que la méthode du maximum de vraisemblance donne encore des estimateurs asymptotiquement efficaces. La technique principale qu'on emploie ici est de réduire convenablement le problème associé à un processus à temps discret, puis d'appliquer les résultats obtenus précédemment. Nous allons indiquer les techniques employées par Billingsley [13] et Bui [17] ainsi qu'une autre qui nous permettent de faire une estimation du paramètre par la méthode du maximum de vraisemblance.

Soit $(\Omega, \mathcal{B}, Pr)$ un espace probabilisé sur lequel est défini un processus de Markov strictement stationnaire à temps continu $\{X_t, t \in T\}$ où T est un intervalle de $\mathbf{R}^+$ et où l'espace des états est un espace mesurable $(\mathcal{X}, \mathcal{S})$, $\mathcal{X}$ étant un ensemble mesurable de $\mathbf{R}^m$, $\mathcal{S}$ la tribu trace de la tribu borélienne de $\mathbf{R}^m$. Pr est en fait supposé dépendre d'un paramètre $\theta \in \Theta$ de sorte que les probabilités de transition stationnaires $P(t, x, A ; \theta)$ dépendent aussi de ce paramètre θ .

(A) Sous l'hypothèse suivante, Billingsley [13] a démontré que le problème de l'estimation de θ du processus donné peut se réduire à un problème d'estimation de θ d'un processus inclus (Imbedded process).

Hypothèse A. — Chaque réalisation est une fonction étagée continue à droite. Pour $\theta \in \Theta$

$$\lim_{t \to 0} P(t, \xi, \{\xi\} ; \theta) = 1$$

uniformément en ξ tel qu'il existe un couple de fonctions $\{ q(.\ ;\theta), q(.\ ,\ .\ ;\theta) \}$ appelé couple de q-fonctions standard (standard pair of q-functions - Doob ch.VI) où $q(.\ ;\theta)$ est bornée. Pour chaque $\theta \in \Theta$

$$\inf_{\xi \in \mathcal{X}} \quad q(\xi ; \theta) > 0$$

Enfin, $P(t, \xi, .\ ; \theta)$ est une fonction $\mathcal{S}$-mesurable.

Nous rappelons que $P(t\ ; \xi, .\ ; \theta)$ est la mesure de transition stationnaire donnée par

$$P(t, \xi, A\ ; \theta) = \Pr \{ X_{t+s} \in A \mid X_s = \xi\ ; \theta \}\ ,\ \forall A \in \mathcal{S}$$

Sous l'thèse ci-dessus, il existe un processus de Markov inclus $\{ Z_n, \rho_n, n \in \mathbf{N}^* \}$ à valeurs dans l'espace des états $\mathcal{X} \times \mathbf{R}$, lié au processus $\{ X_t, t \in T \}$. Les probabilités de transition de ce processus sont données par

$$(4.3.30) \qquad \Pr \{ Z_{n+1} \in A,\ \rho_{n+1} \in B \mid Z_n = \xi, \rho_n = \alpha\ ; \theta \} =$$

$$= \int_A \pi(\xi, d\eta\ ; \theta) \int_B q(\eta\ ; \theta)\ e^{-q(\eta\ ;\ \theta)\beta}\, d\beta$$

où $\pi(\xi, A\ ; \theta) = \dfrac{q(\xi, A\ ; \theta}{q(\xi\ ; \theta)}\ , \qquad \forall \xi \in \mathcal{X}$

La vraisemblance pour une réalisation $\{ X_\tau(\omega)\ 0 \leqslant \tau \leqslant t \}$ est la dérivée de Radon-Nikodym de la restriction de $\Pr(.\ ; \theta)$ sur $\mathcal{S}_t$ tribu borélienne engendrée par $\{ X_\tau\ 0 \leqslant \tau \leqslant t \}$, par rapport à une mesure σ-finie sur $\mathcal{B}$ indépendante de θ . Nous pouvons prendre cette mesure comme la restriction sur $\mathcal{S}_t$ de $\Pr(.\ ; \theta_o)$ où $\theta_o \in \Theta$ est fixé mais arbitraire (ceci n'est pas une restriction comme on peut le voir par le théorème de Halmos-Savage [25]). Ainsi, la dérivée de Radon-Nikodym de la restriction sur $\mathcal{S}_t$ de $\Pr(.\ ; \theta)$ par rapport à la restriction sur $\mathcal{S}_t$ de $\Pr(.\ ; \theta_o)$ est donnée par

$$\delta(Z_1 ; \theta) \prod_{k=1}^{\nu(t)} \left[\frac{e^{-q(Z_k ; \theta)\rho_k}}{e^{-q(Z_k ; \theta_o)\rho_k}} \cdot \delta(Z_k, Z_{k+1} ; \theta) \right].$$

(4.3.31)

$$\cdot \frac{\exp\left[-q(Z_{\nu(t)+1} ; \theta)\left(t - \sum_{k=1}^{\nu(t)} \rho_k\right)\right]}{\exp\left[-q(Z_{\nu(t)+1} ; \theta_o)\left(t - \sum_{k=1}^{\nu(t)} \rho_k\right)\right]}$$

où $\delta(. ; \theta)$ est la dérivée de Radon-Nikodym de la loi initiale pour θ par rapport à la loi initiale pour θ_o en supposant qu'elles sont absolument continues l'une par rapport à l'autre, et

$$\delta(\xi, \eta ; \theta) = \frac{d\, q(\xi, . ; \theta)}{d\, q(\xi, . ; \theta_o)} ,$$

calculée en η .

Maintenant, supposons qu'il existe une mesure σ-finie, λ sur $\mathcal{S}$ par rapport à laquelle $\pi(\xi, . ; \theta)$ est absolument continue de densité $s(\xi, \eta ; \theta)$. Alors, après quelques ajustements de la fonction de vraisemblance qui n'ont d'ailleurs pas de conséquences importantes lorsqu'il s'agit de grands échantillons, le logarithme de vraisemblance peut s'écrire

$$(4.3.32) \quad L_t(\theta) = \sum_{k=1}^{\nu(t)-1} [\log s(Z_k, Z_{k+1} ; \theta) + \log q(Z_{k+1} ; \theta) -$$

$$- \rho_{k+1}\, q(Z_{k+1} ; \theta)] - \sum_{k=1}^{\nu(t)-1} [\log s(Z_k, Z_{k+1} ; \theta_o) +$$

$$+ \log q(Z_{k+1} ; \theta_o) - \rho_{k+1}\, q(Z_{k+1} ; \theta_o)] .$$

Comme θ_o est fixé, en considérant la méthode du maximum de vraisemblance, seule la dérivée de $L_t(\theta)$ par rapport à θ intervient. Ainsi on peut écrire l'équation de vraisemblance

(4.3.33)

$$\frac{\partial L_t(\theta)}{\partial \theta} = 0 = \frac{\partial}{\partial \theta} \left\{ \sum_{k=1}^{\nu(t)-1} \log s(Z_k, Z_{k+1}; \theta) + \log q(Z_{k+1}; \theta) - \right.$$

$$\left. - \rho_{k+1}\, q(Z_{k+1}; \theta) \right\} .$$

Cependant, la densité de la probabilité de transition stationnaire du processus $\{ Z_n, \rho_n,\ n \in \mathbf{N}^* \}$ sur $\mathcal{S} \otimes \mathcal{G}$ est donnée par

(4.3.34) $F(\xi, \eta, \beta; \theta) = s(\xi, \eta; \theta)\, q(\eta; \theta).\ \exp(-\beta\, q(\eta; \theta))$.

Donc, l'équation de vraisemblance d'un échantillon de taille $\nu(t)$ du processus inclus (après avoir supprimé le premier terme) est donnée par (4.3.33).

Ainsi, le problème de l'estimation de θ du processus original à partir d'un échantillon $\{ X_\tau(\omega),\ 0 \leqslant \tau \leqslant t \}$ est équivalent au problème de l'estimation du θ du processus inclus $\{ Z_n, \rho_n, n \in \mathbf{N}^* \}$ pour t assez grand pourvu que les processus satisfassent aux hypothèses du paragraphe 4.3.1.

Donc, nous avons la

PROPOSITION 4.3.4. — *Si* $\{ X_t, t \in T \}$ *est un processus de Markov à temps continu satisfaisant à l'hypothèse* (A) *de ce paragraphe et si le processus inclus satisfait aux hypothèses du paragraphe* 4.3.1., *l'équation de vraisemblance donnée par* (4.3.33) *a, avec une probabilité qui tend vers* 1 *quand* $n \to \infty$, *une racine* $\hat{\theta}_t(\underset{\sim}{X}^{(t)})$ *qui est un estimateur convergent et asymptotiquement efficace de la vraie valeur du paramètre* θ .

(B) Méthode de Bui [17] : Dans le même cas que ci-dessus, supposons que les hypothèses suivantes sont satisfaites

Hypothèses B. — Il existe un $\theta_1 \in \Theta$ tel que

$$\lim_{t \to 0} P(t, \xi, \{\xi\}; \theta_1) = 1$$

uniformément en $\xi \in \mathcal{X}$; $\dfrac{\partial P(t, \xi, A; \theta)}{\partial \theta}$

existe et est continue en $\theta \in \Theta$ pour $\forall t \in T$, $\forall \xi \in \mathcal{X}$ et $\forall A \in \mathcal{S}$.

$$\lim_{t \to 0} \frac{1}{t} P(t, \xi, A ; \theta)$$

existe uniformément en $\theta \in \Theta$, $\forall \xi \in \mathcal{X}$, $\forall A \in \mathcal{S}$

Il existe une mesure μ sur $\mathcal{S}$, indépendante de θ telle que $P(t, \xi, . ; \theta) \ll \mu \ \forall \theta \in \Theta$. De plus, la condition de Doeblin est satisfaite (Doob ch.VI) et $\forall \theta \in \Theta$, $\forall A \in \mathcal{S}$,

$$\lim_{t \to T} P(t, \xi, A ; \theta) = \pi(A ; \theta)$$

existe et est indépendante de ξ. Il s'ensuit que

$\pi(. ; \theta) \ll \mu$ et, par conséquent, la dérivée de Radon-Nikodym $\tilde{\pi}(x ; \theta)$ de $\pi(. ; \theta)$ par rapport à μ existe.
Enfin, on suppose que

$$\frac{\partial^\nu q(x, . ; \theta)}{\partial \theta^\nu}, \quad \frac{\partial^\nu \tilde{\pi}(x ; \theta)}{\partial \theta^\nu} \quad \text{et} \quad \frac{\partial^\nu \tilde{q}(x, y ; \theta)}{\partial \theta^\nu}$$

pour $\nu = 1, 2, 3$ existent et sont bornés, où $\tilde{q}(x, y ; \theta)$ est la densité de $q(x, . ; \theta)$ par rapport à μ.

Supposons de plus que

$$\beta_\nu(x) = \operatorname*{supess}_{\theta \in \Theta} \left| \frac{\partial^\nu \tilde{\pi}(x ; \theta)}{\partial \theta^\nu} \right|$$

et

$$\gamma_\nu(x, y) = \operatorname*{supess}_{\theta \in \Theta} \left| \frac{\partial^\nu \tilde{q}(x, y ; \theta)}{\partial \theta^\nu} \right|$$

sont intégrables $-\mu$, la dernière expression l'étant pour $\forall x \in \mathcal{X}$

(Rappelons que $q(x, . ; \theta)$ est la même mesure que dans le cas (A)).

Considérons une réalisation $\{X_t(\omega),\ 0 \leqslant t \leqslant t_o\}$ du processus. Notons $\tau_{t_o}(A\ ;\ \omega)$ la durée de séjour en $A \in \mathcal{S}$ et $N_{t_o}(A \times B\ ;\ \omega)$ la fréquence de passage de $A \in \mathcal{S}$ à $B \in \mathcal{S}$ dans l'observation correspondant à $\omega \in \Omega$ et à $[0, t_o]$.

Alors l'équation du maximum de vraisemblance s'écrit

$$(4.3.35)\quad \frac{1}{t_o}\ \frac{\partial L_t(\theta)}{\partial \theta} = \frac{1}{t_o}\ \frac{\partial}{\partial \theta} \log \tilde{\pi}\ (X(0,\omega)\ ;\ \theta) -$$

$$- \frac{1}{t_o} \int_{\mathcal{X}} q(x,\theta)\, d\tau_{t_o}(x\ ;\ \omega) + \frac{1}{t_o} \int_{\mathcal{X}\times\mathcal{X}} \frac{\partial}{\partial \theta} \log \tilde{q}(x,y;\theta)\ .$$

$$.\ dN_{t_o}(x,y\ ;\ \theta) = 0$$

Ainsi, nous avons la

PROPOSITION 4.3.4.bis. — *L'équation de vraisemblance* (4.3.33) *a une racine* $\hat{\theta}_t$ *pour* $t \geqslant t_o(\omega)$ *où* $t_o(\omega)$ *est une variable presque sûrement finie.* $\hat{\theta}_t$ *est un estimateur convergent et asymptotiquement efficace de la vraie valeur de* θ .

(C) Sous les hypothèses de la méthode (A), le théorème suivant dû à Adhikari [1] démontre que, si les observations sont faites sur un ensemble dénombrable partout dense de (0, t) plutôt que sur toute la trajectoire, la fonction de vraisemblance est encore la même que dans le cas (A) et par conséquent, la méthode s'applique encore.

THÉORÈME D'ADHIKARI . — *Soit* $\{X_t, t \in T\}$ *où* T *est un intervalle sur la droite réelle, un processus à valeurs réelles défini sur* $(\Omega, \mathcal{B})$ *où* $\mathcal{B}$ *est la plus petite tribu sur* Ω *par rapport à laquelle toutes les* X_t, $t \in T$ *sont mesurables. Supposons que* μ_1 *et* μ_2 *sont deux mesures de probabilité sur* $\mathcal{B}$, *telles que* $\mu_1 \ll \mu_2$. *Si le processus* $\{X_t, t \in T\}$ *est tel que*

$\lim_{\tau \to t} X_\tau = X_t$ *presque partout* $[\mu_2]$ *pour tout* $t \in T$ *et si* $\{t_1\}$ *est une suite dénombrable dense dans* T, *alors, on a*

$$\frac{d\mu_1}{d\mu_2} = \lim_{n \to \infty} \frac{f_{\mu_1}^{(t_1,\ldots,t_n)}(X_{t_1},\ldots,X_{t_n})}{f_{\mu_2}^{(t_1,\ldots,t_n)}(X_{t_1},\ldots,X_{t_n})}$$ *presque sûrement* $[\mu_2]$ *en supposant que toutes les lois de probabilité à nombre fini de variables possèdent des densités par rapport aux deux mesures.*

De ce théorème, on déduit le résultat suivant

Soit $\mathfrak{M}$ une classe de mesures de probabilité sur $\mathcal{B}$ et λ une probabilité sur $\mathcal{B}$ telle que $\mathfrak{M} \equiv \lambda$.

Soit $\{t_1\}$ un sous-ensemble dénombrable dense de T . Désignons par $F(\omega)$ l'application $F(\omega) = \{X_t(\omega),\ t \in T\}$ de $(\Omega, \mathcal{B})$ dans $(\mathbf{R}^\infty, \mathcal{C}^\infty)$; alors $F(\omega)$ est un résumé exhaustif pour la classe $\mathfrak{M}$.

Remarque. — Dans le cas des processus à temps continu, comme dans celui des processus à temps discret, les résultats précédents s'appliquent en particulier lorsque l'espace des états ne contient qu'un nombre fini de points.

4.4.1. Processus de Poisson

Considérons un processus stochastique $\{X(t),\ t \geqslant 0\}$ défini sur $(\Omega, \mathcal{B})$ avec l'espace des états $\mathbf{N}$ satisfaisant les propriétés suivantes

(1) $X(t, \omega)$ est le nombre de fois où un événement se produit durant l'intervalle de temps $[0, t]$.

(2) Le nombre de fois qu'un événement se produit dans tout intervalle fini est fini.

(3) Le processus $\{X(t),\ t \geqslant 0\}$ est à accroissements indépendants.

Puisqu'on compte à partir du moment $t = 0$, on suppose que $X(0, \omega) \equiv 0$, $\forall \omega$, .

Alors, il est bien connu [Doob ch.VII] que tout processus satisfaisant aux conditions ci-dessus est un processus de Poisson dont la probabilité est donnée par

$$P \{ X(t,\omega) - X(s,\omega) = m \}$$

$$(4.4.1.) \quad = \{ \exp [\psi(s) - \psi(t)] \} . \frac{[\psi(s) - \psi(t)]^m}{m !}$$

pour $s < t$

où $\psi(t) = \log \Phi(t)$

$$\Phi(t) = P \{ X(t,\omega) - X(0,\omega) = 0 \} , \quad t \geqslant 0$$

(Doob. ch.VII.4) .

Maintenant, si l'on suppose de plus que les accroissements indépendants sont stationnaires, alors la forme de $\Phi(t)$ est déterminée uniquement comme

$$\Phi(t) = e^{-\theta t} ,$$

pour une constante $\theta > 0$. θ est appelé "intensité". Maintenant se pose le problème d'estimer la vraie valeur du paramètre θ à partir d'une suite d'observations du processus dans $[0, t]$.

Supposons que nous avons une observation du processus dans l'intervalle $[0, t]$. Soit $N(t)$ le nombre d'événements observés durant ce temps. Alors, la vraisemblance de $N(t)$ est donnée par

$$(4.4.2.) \quad P \{ X(t, \omega) - X(0, \omega) = N(t) \} =$$

$$= e^{-\theta t} \frac{(\theta t)^{N(t)}}{N(t)!} = L(t, \omega ; \theta) ,$$

qui n'est rien d'autre que la loi de Poisson.

Donc, l'équation de vraisemblance est

$$(4.4.3.)\quad \frac{d}{d\theta}\ \log L(t,\omega;\theta) = \frac{d}{d\theta}\ [-\theta t + N(t)\log \theta t - \text{Log } N(t)\,!\,] =$$

$$= \frac{N(t)}{\theta} - t = 0\ .$$

Or, la racine unique de cette équation est

$$\hat{\theta}_t(\omega) = \frac{N(t,\omega)}{t}\quad ,$$

Pour démontrer que $\hat{\theta}_t$ est asymptotiquement efficace, on a

$$E_\theta\left(\left[\frac{\partial \log L}{\partial \theta}\right)^2\right] = \frac{t}{\theta} = I(t,\theta)$$

Donc

$$\frac{1}{\sqrt{I(t,\theta_\circ)}}\ \left(\frac{\partial \log L}{\partial \theta}\right)_{\theta=\theta_\circ} = \sqrt{\frac{\theta_\circ}{t}}\left(\frac{N(t)-t\theta_\circ}{\theta_\circ}\right) = \frac{(N(t)-t\theta_\circ)}{\sqrt{t\theta_\circ}}$$

et $\sqrt{I(t,\theta_\circ)}\ (\hat{\theta}_t - \theta_\circ) = \sqrt{\frac{t}{\theta_\circ}}\left(\frac{N(t)-t\theta_\circ}{t}\right) = \frac{(N(t)-t\theta_\circ)}{\sqrt{t\theta_\circ}}\quad ,$

ce qui démontre le résultat.

4.4.2. Processus du mouvement brownien (Processus de Wiener-Lévy) .

Nous démontrons que la méthode du maximum de vraisemblance donne des estimateurs asymptotiquement efficaces dans le cas du processus de Wiener-Lévy.

Considérons le processus $\{X(t),\ t \geqslant 0\}$ défini sur $(\Omega, \mathcal{B}^{(\mathcal{S})})$ l'espace des états étant $(\mathcal{X}, \mathcal{S})$ où $\mathcal{X}$ est un ensemble borélien de $\mathbf{R}$ et $\mathcal{S}$ la tribu trace de la tribu borélienne de $\mathbf{R}$ et caractérisé comme suit

(1) pour chaque couple $s,t \geqslant 0$, la différence $X(t)-X(s)$ est une variable aléatoire qui suit une loi gaussienne d'espérance mathématique 0 et de variance

$$\mathrm{Var}(X(t) - X(s)) = \sigma^2 \mid t-s \mid$$

(2) pour chaque suite finie de point $0 = t_o < t_1 < \ldots < t_n$, les variables aléatoires $X(t_1) - X(t_o),\ldots,X(t_n) - X(t_{n-1})$, sont indépendantes.

Le processus décrit ici est appelé processus à accroissements indépendantes stationnaires (Doob, ch.VIII).

En général, on suppose aussi que $X(0) \equiv 0$ presque - sûrement, et que $E(X(t)) = 0, \forall\, t > 0$.

Dans ces conditions, si $t_1,\ldots,t_n$ sont n instants et $X(t_1, \omega),\ldots,$ $,\ldots, X(t_n, \omega)$ est une observation du processus, les variables $X(t_1) -$ $-X(t_o),\ldots,X(t_n) - X(t_{n-1})$ suivent une loi gaussienne, dont la densité est donnée par

$$(4.4.4.) \qquad f_n(\underset{\sim}{X};\, \sigma^2) = \frac{1}{\sigma^n (2\pi)^{n/2}} \prod_{i=1}^{n} \frac{1}{(t_i - t_{i-1})^{1/2}} \cdot$$

$$\cdot \exp\left\{ \frac{-(X(t_i) - X(t_{i-1}))^2}{2\sigma^2 (t_i - t_{i-1})} \right\} =$$

$$= \frac{1}{\sigma^n (2\pi)^{1/2}} \left[\exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \frac{(X(t_i) - X(t_{i-1}))^2}{(t_i - t_{i-1})} \right\} \right] \cdot$$

$$\cdot \prod_{i=1}^{n} \frac{1}{(t_i - t_{i-1})^{1/2}} \cdot$$

Maintenant, supposons qu'on veut faire une estimation de $\sigma^2 = \theta$ à partir d'un échantillon de taille n du processus, c'est-à-dire à partir des observations des n instants $0 < t_1 < \ldots < t_n$. L'équation de vraisemblance s'écrit

(4.4.5.) $$\frac{\partial \log f_n}{\partial \theta} = \frac{\partial}{\partial \theta}\left[-\frac{1}{2\theta} \sum_{i=1}^{n} \frac{(X(t_i) - X(t_{i-1}))^2}{(t_i - t_{i-1})} - \right.$$

$$\left. -\frac{n}{2} \log \theta - \frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{1}^{n} \log (t_i - t_{i-1}) \right] = 0$$

c'est-à-dire $$\frac{1}{2\theta^2} \sum_{i=1}^{n} \frac{(X(t_i) - X(t_{i-1}))^2}{(t_i - t_{i-1})} - \frac{n}{2\theta} = 0 .$$

Or $$\hat{\theta}_n(\omega) = \frac{1}{n} \sum_{i=1}^{n} \frac{(X(t_i, \omega) - X(t_{i-1}, \omega))^2}{(t_i - t_{i-1})}$$

(4.4.6.)

On voit facilement que $\hat{\theta}_n$ est asymptotiquement efficace, car

$$I(n, \theta) = E_o\left[\left(\frac{\log f_n}{\theta} \right)^2 \right] = \frac{n}{2\theta^2}$$

et

$$\frac{1}{\sqrt{I(n, \theta_o)}} \left(\frac{\partial \log f_n}{\partial \theta} \right)_{\theta = \theta_o} = \left[\frac{1}{\sqrt{2n, \theta_o}} \sum_{i=1}^{n} \frac{(X(t_i) - X(t_{i-1}))^2}{(t_i - t_{i-1})} - \sqrt{\frac{n}{2}} \right]$$

$$\sqrt{I(n, \theta_o)} \ (\hat{\theta}_n - \theta_o) = \left[\frac{1}{\sqrt{2n, \theta_o}} \sum_{i=1}^{n} \frac{(X(t_i) - X(t_{i-1}))^2}{t_i - t_{i-1}} - \sqrt{\frac{n}{2}} \right]$$

d'où le résultat.

5. Estimation pour des processus stochastiques (cas de plusieurs paramètres)

5.0. Introduction.

Dans ce chapitre, nous formulons la définition de l'efficacité asymptotique dans le cas de plusieurs paramètres comme une généralisation naturelle de celle introduite dans le cas d'un paramètre.

Comme dans le chapitre précédent, la définition est basée sur l'utilité d'un estimateur au lieu d'un échantillon pour les problèmes des tests d'hypothèses concernant des paramètres. Donc la définition de l'efficacité asymptotique est formulée en rapport avec les dérivées partielles du logarithme de vraisemblance. On montre que la méthode du maximum de vraisemblance donne des estimateurs convergents et asymptotiquement efficaces pour les vraies valeurs des paramètres.

5.1. Préliminaires.

Nous nous plaçons dans le cas du paragraphe 4.1. Supposons que les probabilités de dimensions finies des variables $\{ P^{(n)}(. \ ; \theta) \}_{n \in \mathbf{N}^*}$ et la probabilité stationnaire $P(. \ ; \theta)$ sont connues sauf pour la valeur d'un paramètre vectoriel $\underset{\sim}{\theta} \in \Theta$, ouvert de $\mathbf{R}^k$. Notons $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)$. Nous allons faire des hypothèses analogues à celles du paragraphe 4.1.

Les hypothèses 5.1. et 5.2. sont les mêmes qu'en 4.1. et 4.2.

Hypothèse 5.3. — Les dérivées partielles $\dfrac{\partial f_n}{\partial \theta_i} \cdot \dfrac{\partial^2 f_n}{\partial \theta_i \, \partial \theta_j}$ existent presque partout $[\lambda^{(n)}]$ pour $i, j = 1, .., k$, pour tout $\underset{\sim}{\theta} \in \Theta$ et sont continues en $\underset{\sim}{\theta} \in \Theta$. Aussi

$$\frac{\partial}{\partial \theta_i} \int_{A^{(n)}} f_n \, d\lambda^{(n)} = \int_{A^{(n)}} \frac{\partial f_n}{\partial \theta_i} \, d\lambda^{(n)}$$

$$\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\int_{A^{(n)}} f_n\, d\lambda^{(n)} = \int_{A^{(n)}} \frac{\partial^2 f_n}{\partial\theta_i\,\partial\theta_j}\, d\lambda^{(n)}$$

$\forall A^{(n)} \in \mathcal{S}^{(n)}$, $i, j = 1,\ldots,k$, $\forall \underset{\sim}{\theta} \in \Theta$, $\forall n \in \mathbf{N}^*$.

Supposons que l'ensemble minimal des statistiques pour faire l'estimation de $\underset{\sim}{\theta}$ est de dimension k . Soit $\underset{\sim}{T}_n(\underset{\sim}{X}^{(n)}) = (T_{1n}(\underset{\sim}{X}^{(n)}),\ldots,T_{kn}(\underset{\sim}{X}^{(n)}))$ une application de l'espace des échantillons dans $(\mathbf{R}^k, \mathcal{C}^k)$ et $\Phi_{\underset{\sim}{T}_n}(.\,;\theta)$ la probabilité induite sur $\mathcal{C}^k$. Alors $\Phi_{T_n}(.\,;\theta)$ est absolument continue par rapport à la mesure image λ' de $\lambda^{(n)}$ par $\underset{\sim}{T}_n$.

Soit $\varphi_n(\underset{\sim}{T}_n\,;\underset{\sim}{\theta})$ la densité par rapport à λ' . Supposons que

$$\frac{\partial}{\partial\theta_i}\int_{A^{(k)}} \varphi_n\, d\lambda' = \int_{A^{(k)}} \frac{\partial\varphi_n}{\partial\theta_i}\, d\lambda'$$

$$\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\int_{A^{(k)}} \varphi_n\, d\lambda' = \int_{A^{(k)}} \frac{\partial^2\varphi_n}{\partial\theta_i\,\partial\theta_j}\, d\lambda' \quad .$$

pour $\forall\, i,j = 1,\ldots,k$, $\forall\, n \in \mathbf{N}^*$, $\forall \underset{\sim}{\theta} \in \Theta$, $\forall A^{(k)} \in \mathcal{C}^k$

Cette hypothèse implique que

$$E_{\underset{\sim}{\theta}}\left(\frac{\partial \log f_n}{\partial\theta_i}\right) = 0$$

$$E_{\theta}\left(\frac{\partial \log f_n}{\partial\theta_i}\cdot\frac{\partial \log f_n}{\partial\theta_j}\right) = -E_{\underset{\sim}{\theta}}\left(\frac{\partial^2 \log f_n}{\partial\theta_i\,\partial\theta_j}\right) = I_{ij}(n\,;\underset{\sim}{\theta})$$

Soient $I(n\,;\underset{\sim}{\theta}) = (I_{ij}(n\,;\underset{\sim}{\theta}))$ la matrice d'information dans un échantillon de taille n et $I(\underset{\sim}{T}_n\,;\underset{\sim}{\theta}) = (I_{ij}(\underset{\sim}{T}_n\,;\underset{\sim}{\theta}))$ la matrice d'information dans la statistique $\underset{\sim}{T}_n$.

Utilisons les notations suivantes

$$Z_{in}(\underset{\sim}{\theta}) = Z_i(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}) = \frac{\partial \log f_n}{\partial \theta_i} , \quad i = 1,\ldots,k$$

$$\underset{\sim}{Z}_n(\underset{\sim}{\theta}) = (Z_{1n}(\underset{\sim}{\theta}),\ldots, Z_{kn}(\underset{\sim}{\theta}))^T$$

$$Y_{in}(\underset{\sim}{\theta}) = Y_i(\underset{\sim}{T}_n ; \underset{\sim}{\theta}) = \frac{\partial \log \varphi_n}{\partial \theta_i} , \quad i = 1,\ldots, k$$

$$\underset{\sim}{Y}_n(\underset{\sim}{\theta}) = (Y_{1n}(\underset{\sim}{\theta}),\ldots, Y_{kn}(\underset{\sim}{\theta}))^T$$

$$U_{in}(\underset{\sim}{\theta}) = (T_{in}(\underset{\sim}{X}^{(n)}) ; \theta_i) , \quad i = 1,\ldots,k$$

$$\underset{\sim}{U}_n(\underset{\sim}{\theta}) = (U_{1n}(\underset{\sim}{\theta}),\ldots, U_{kn}(\underset{\sim}{\theta}))^T .$$

(Notons que les expressions Z_{in}, Y_{in}, et U_{in} ne sont pas réduites ici à la différence du chapitre précédent).

Alors les résultats suivants analogues au cas d'un paramètre sont immédiats.

(Dorénavant, les vecteurs $(\underset{\sim}{.})$ utilisés sont des vecteurs colonnes et leurs transposés seront notés par $(\underset{\sim}{.})^T$) .

(1) $$E_{\underset{\sim}{\theta}}(\underset{\sim}{Z}_n(\underset{\sim}{\theta}) \mid \underset{\sim}{T}_n) = \underset{\sim}{Y}_n(\underset{\sim}{\theta}) \text{ presque partout } [\lambda^{(n)}]$$

(2) $$E_{\underset{\sim}{\theta}}(\underset{\sim}{Y}_n(\underset{\sim}{\theta})\, \underset{\sim}{Z}_n^T(\underset{\sim}{\theta})) = E_{\underset{\sim}{\theta}}(\underset{\sim}{Y}_n(\underset{\sim}{\theta})\, E_{\underset{\sim}{\theta}}(\underset{\sim}{Z}_n^T(\underset{\sim}{\theta}) \mid \underset{\sim}{T}_n))$$
$$= E_{\underset{\sim}{\theta}}(\underset{\sim}{Y}_n(\underset{\sim}{\theta})\, \underset{\sim}{Y}_n^T(\underset{\sim}{\theta})) = I(\underset{\sim}{T}_n ; \underset{\sim}{\theta})$$

(3) Considérons la matrice de covariances de $(\underset{\sim}{Y}_n^T(\underset{\sim}{\theta}), \underset{\sim}{Z}_n^T(\underset{\sim}{\theta}))$ donnée par

$$\begin{bmatrix} I(\underset{\sim}{T}_n ; \underset{\sim}{\theta}) & I(\underset{\sim}{T}_n ; \underset{\sim}{\theta}) \\ I(\underset{\sim}{T}_n ; \underset{\sim}{\theta}) & I(n ; \underset{\sim}{\theta}) \end{bmatrix}$$

qui est semi définie-positive. Puisque $I(n ; \underset{\sim}{\theta})$ est définie positive par l'hypothèse $\forall\, \underset{\sim}{\theta} \in \Theta$,

$$\text{dét } [\, I(\underset{\sim}{T}_n ; \underset{\sim}{\theta}) - I(\underset{\sim}{T}_n ; \underset{\sim}{\theta})\, I^{-1}(n ; \underset{\sim}{\theta})\, I(\underset{\sim}{T}_n ; \underset{\sim}{\theta}) \,] \geqslant 0$$

c'est-à-dire

$$\text{dét } I(\underset{\sim}{T}_n ; \underset{\sim}{\theta})\ \text{dét } [\, I - I^{-1}(n ; \underset{\sim}{\theta})\ I(\underset{\sim}{T}_n ; \underset{\sim}{\theta}) \,] \geqslant 0$$

c'est-à-dire

$$\text{dét } [\, I - I^{-1}(n ; \underset{\sim}{\theta})\ I(\underset{\sim}{T}_n ; \underset{\sim}{\theta}) \,] \geqslant 0$$

puisque $\text{dét } I(\underset{\sim}{T}_n ; \underset{\sim}{\theta}) \geqslant 0$.

c'est-à-dire $\text{dét } I(n ; \underset{\sim}{\theta})\ \text{dét } [\, I(n ; \underset{\sim}{\theta}) - I(\underset{\sim}{T}_n ; \underset{\sim}{\theta}) \,] \geqslant 0$

ou $\text{dét } [\, I(n ; \underset{\sim}{\theta}) - I(\underset{\sim}{T}_n ; \underset{\sim}{\theta}) \,] \geqslant 0$

ce qui démontre que

(5.1.4.) $I(n ; \underset{\sim}{\theta}) - I(\underset{\sim}{T}_n ; \underset{\sim}{\theta})$

est semi-définie-positive.

(4) Supposons que la loi asymptotique de $(\underset{\sim}{Z}_n^T(\underset{\sim}{\theta}), \underset{\sim}{U}_n^T(\underset{\sim}{\theta}))$ existe avec deux moments finis. Soit $\underset{\sim}{c}$ un point de continuité de la loi asymptotique de $\underset{\sim}{U}_n$ et soit

$$\underset{\sim}{U}_n^{\underset{\sim}{c}} = \underset{\sim}{U}_n \quad \text{si} \quad |\, \underset{\sim}{U}_n \,| < \underset{\sim}{c}$$

$$0 \quad \text{sinon} .$$

(Ici la valeur absolue du vecteur est entendue comme la valeur absolue de chaque composante).

(5.1.5.) Alors $\liminf_{n \to \infty} \dfrac{I_{ii}(\underset{\sim}{T}_n ; \underset{\sim}{\theta})}{I_{ii}(n ; \underset{\sim}{\theta})} \geqslant \rho_{ij}^{\underset{\sim}{c}}$

où ρ_{ij} est la corrélation asymptotique de $Z_{in}(\underset{\sim}{\theta})$ et $U_{jn}(\underset{\sim}{\theta})$, $\forall\, i, j = 1, \ldots, k$.

Ceci est immédiat en observant que

$$\mathrm{Var}_{\underset{\sim}{\theta}}(U^{c}_{\underset{\sim}{jn}}(\underset{\sim}{\theta})\ \mathrm{Var}_{\underset{\sim}{\theta}}(Y_{in}(\underset{\sim}{\theta}))) \geqslant \left| E_{\underset{\sim}{\theta}}(U^{c}_{\underset{\sim}{jn}}(\underset{\sim}{\theta})\, Y_{in}(\underset{\sim}{\theta})) \right|^2 =$$

$$= (\alpha^{c}_{\underset{\sim}{ijn}})^2 = \left| E_{\underset{\sim}{\theta}}(U^{c}_{\underset{\sim}{jn}}(\underset{\sim}{\theta})\ Z_{in}(\underset{\sim}{\theta}) \right|^2$$

et par conséquent ,

$$\liminf_{n \to \infty} \frac{I_{ii}(\underset{\sim}{T}_n ; \underset{\sim}{\theta})}{I_{ii}(n ; \underset{\sim}{\theta})} \geqslant \rho^2_{ij}$$

(5) Si la corrélation asymptotique de Z_{in} et U_{in} est l'unité, alors la corrélation asymptotique de Z_{in} et Y_{in} est aussi l'unité.

$$E_{\underset{\sim}{\theta}}(Y_{in}(\underset{\sim}{\theta})\ Z_{in}(\underset{\sim}{\theta})) = E_{\underset{\sim}{\theta}}((Y_{in}(\underset{\sim}{\theta}))^2 = I_{ii}(\underset{\sim}{T}_n ; \underset{\sim}{\theta})$$

Donc $\quad I^2_{ii}(\underset{\sim}{T}_n ; \underset{\sim}{\theta}) \leqslant I_{ii}(n ; \underset{\sim}{\theta})\ I_{ii}(\underset{\sim}{T}_n ; \underset{\sim}{\theta})$

ou

$$(5.1.6.)\quad \rho^2_{ii} \leqslant \limsup_{n \to \infty} \frac{I_{ii}(\underset{\sim}{T}_n ; \underset{\sim}{\theta})}{I_{ii}(n ; \underset{\sim}{\theta})} \leqslant 1$$

d'où le résultat.

5. 2. Comme extension naturelle du cas d'un paramètre et d'après les propriétés optimales des tests basés sur les dérivées partielles du logarithme de vraisemblance, nous formulons la définition suivante

DÉFINITION 5.2.1. – Un estimateur $\underset{\sim}{T}_n(\underset{\sim}{X}^{(n)})$ est dit asymptotiquement efficace pour $\underset{\sim}{\theta}_o$ si

$$\left| A(n ; \underset{\sim}{\theta}_o)\, \underset{\sim}{Z}_n(\underset{\sim}{\theta}_o) - \alpha(n ; \underset{\sim}{\theta}_o) - \beta(n ; \underset{\sim}{\theta}_o)\, \underset{\sim}{U}_n(\underset{\sim}{\theta}_o) \right|$$

tend vers 0 en probabilité où $A(n ; \underset{\sim}{\theta})$ et $\beta(n ; \underset{\sim}{\theta})$ sont des matrices non nulles d'ordre $(k \times k)$ qui ne dépendent pas des réalisations du processus

mais peuvent dépendre de n et $\underset{\sim}{\theta}$. En outre, $A(n ; \underset{\sim}{\theta})$ et $\beta(n ; \underset{\sim}{\theta})$ ne tendent pas vers les matrices dégénérées lorsque $n \to \infty$, $\forall \underset{\sim}{\theta} \in \Theta$. $\alpha(n ; \underset{\sim}{\theta})$ est aussi une matrice d'ordre $(k \times 1)$ qui ne dépend pas des réalisations du processus, mais peut dépendre de n et $\underset{\sim}{\theta}$. C'est naturel de faire l'hypothèse que

$$\text{ni} \quad A(n ; \underset{\sim}{\theta}_o)\, \underset{\sim}{Z}_n(\underset{\sim}{\theta}_o) \qquad \text{ni} \quad \beta(n ; \underset{\sim}{\theta}_o)\, \underset{\sim}{U}_n(\underset{\sim}{\theta}_o)$$

ne tendent vers 0 en probabilité lorsque $n \to \infty$. Sinon, on peut toujours trouver des matrices $A(n ; \underset{\sim}{\theta}_o)$ et $\beta(n ; \underset{\sim}{\theta}_o)$ satisfaisant à la condition (5.2.1.) .

Nous faisons les hypothèses supplémentaires suivantes

Hypothèse 5.4. — Les dérivées partielles $\dfrac{\partial^3 \log f_n}{\partial\theta_i\, \partial\theta_j\, \partial\theta_l}$ existent pour $\forall\, i, j, l = 1,\ldots,k$, $\forall\, n \in \mathbf{N}^*$ et $\forall\, \underset{\sim}{\theta} \in \Theta$ et il existe une fonction $H_n(\underset{\sim}{X}^{(n)})$ telle que

$$(5.2.2.) \qquad \left| \frac{\partial^3 \log f_n}{\partial\theta_i\, \partial\theta_i\, \partial\theta_l} \right| < H_n(\underset{\sim}{X}^{(n)}) \qquad \begin{matrix} \forall\, \theta \in \Theta \\ \forall\, i, j, l = 1,\ldots,k \end{matrix}$$

et telle que

$$(5.2.3.) \qquad \frac{E_\theta(\,|\, H_n(\underset{\sim}{X}^{(n)})\,|\,)}{I_{max}(n ; \underset{\sim}{\theta})} \leqslant M < \infty \qquad \forall\, n \in \mathbf{N}^*$$

où $I_{max}(n ; \underset{\sim}{\theta}) = \underset{i}{\text{Max}}\ I_{ii}(n ; \underset{\sim}{\theta}) \to \infty$ quand $n \to \infty$.

Hypothèse 5.5. — Pour $\forall\, \underset{\sim}{\theta} \in \Theta$ et $i, j, = 1,\ldots,k$

$$(5.2.4.) \qquad \lim_{n \to \infty} \cdot\ \mathrm{Var}_\theta \left[\frac{1}{I_{max}(n ; \underset{\sim}{\theta})} \frac{\partial^2 \log f_n}{\partial\theta_i\, \partial\theta_j} \right] = 0$$

Hypothèse 5.6. — Pour $\forall\, \underset{\sim}{\theta} \in \Theta$

$$(5.2.5.) \qquad \lim_{n \to \infty} (I_{max}(n ; \underset{\sim}{\theta}))^{-1}\, I(n ; \underset{\sim}{\theta})$$

existe et est non dégénérée.

Remarque. – Les hypothèses semblent très restrictives mathématiquement, mais elles sont en général remplies dans la plupart des cas intéressants et on peut les affaiblir dans les cas particuliers.

PROPOSITION 5.2.1. – *Sous les hypothèses* 5.1. - 5.6. *l'équation de vraisemblance a, avec une probabilité qui tend vers* 1 *quand* $n \to \infty$ *, une racine* $\hat{\theta}_n(\underset{\sim}{X}^{(n)})$ *qui est un estimateur convergent et asymptotiquement efficace de la vraie valeur du paramètre* $\underset{\sim}{\theta}$ *(soit* $\underset{\sim}{\theta}_o$ *)* .

DÉMONSTRATION . – La démonstration se fait essentiellement comme dans le cas classique et nous donnerons les grandes lignes de la démonstration.

Ecrivons

(5.2.6.)

$$\frac{\partial \log f_n(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta})}{\partial \theta_i} = \left(\frac{\partial \log f_n}{\partial \theta_i}\right)_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_o} + (\underset{\sim}{\theta} - \underset{\sim}{\theta}_o)^T \left(\frac{\partial^2 \log f_n}{\partial \theta_i \partial \theta_j}\right)_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_o} +$$

$$+ \frac{1}{2} (\underset{\sim}{\theta} - \underset{\sim}{\theta}_o)^T \left(\left(\frac{\partial^3 \log f_n}{\partial \theta_i \partial \theta_j \partial \theta_l} \right)_{\underset{\sim}{\theta}=\underset{\sim}{\theta}^{i*}} \right) \cdot (\underset{\sim}{\theta} - \underset{\sim}{\theta}_o)$$

pour $i = 1,\ldots,k$ ou $j, l = 1,\ldots,k$

et $\| \underset{\sim}{\theta} - \underset{\sim}{\theta}^{i*} \| < \| \underset{\sim}{\theta} - \underset{\sim}{\theta}_o \|$.

Suivant la notation matricielle de [3] , écrivons le système des équations ci-dessus comme suit

$$\underset{\sim}{l}(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}_o) + (M(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}_o)) (\underset{\sim}{\theta} - \underset{\sim}{\theta}_o) + V^{(1)}(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}, \underset{\sim}{\theta}_o) = 0$$

(5.2.7.)

où $\underset{\sim}{l}(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta})$ est le vecteur colonne de composantes

$$\left(\frac{\partial \log f_n}{\partial \theta_i}\right) \quad i = 1,\ldots,k \ .$$

$M(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta})$ est la matrice dont le (i, j)ème élément est $\dfrac{\partial^2 \log f_n}{\partial\theta_i \, \partial\theta_j}$

i, j = 1,...,k et $V^{(1)}(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}, \underset{\sim}{\theta}_o)$ est le vecteur colonne de composantes $\frac{1}{2} (\underset{\sim}{\theta} - \underset{\sim}{\theta}_o)^T \; S_1(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}, \underset{\sim}{\theta}_o) \; (\underset{\sim}{\theta} - \underset{\sim}{\theta}_o)$

$$\text{Si } (.) = \left| \left(\frac{\partial^3 \log f_n}{\partial\theta_i \, \partial\theta_j \, \partial\theta_l} \right)_{\underset{\sim}{\theta} = \underset{\sim}{\theta}^{i*}} \right| \quad j, l = 1,\ldots,k$$

pour i = 1,...,k et $\| \underset{\sim}{\theta}^{i*} - \underset{\sim}{\theta}_o \| < \| \underset{\sim}{\theta} - \underset{\sim}{\theta}_o \|$

où $\| . \|$ est la norme euclidienne.

Par prémultiplication par $I^{-1}(n ; \underset{\sim}{\theta}_o)$

$$(5.2.8.) \quad I^{-1}(n ; \underset{\sim}{\theta}_o) \, \underset{\sim}{l}(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}_o) + I^{-1}(n ; \underset{\sim}{\theta}_o) \, M(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}_o) \, (\underset{\sim}{\theta} - \underset{\sim}{\theta}_o) +$$

$$+ \; I^{-1}(n ; \underset{\sim}{\theta}_o) \, V^{(1)}(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}, \underset{\sim}{\theta}_o) = 0$$

Puisque $E_{\underset{\sim}{\theta}_o} (I^{-1}(n ; \underset{\sim}{\theta}_o) \, \underset{\sim}{l}(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}_o)) = 0$ et par les hypothèses 5.4. et 5.6., on a $I^{-1}(n ; \underset{\sim}{\theta}_o) \, \underset{\sim}{l}(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}_o) \to 0$ en probabilité. D'après les hypothèses 5.4., 5.5. et 5.6., puisque

$$E_{\underset{\sim}{\theta}_o} [I^{-1}(n ; \underset{\sim}{\theta}_o) \, M(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}_o)] = - I_k ,$$

matrice identité d'ordre $(k \times k)$, on a

$$I^{-1}(n ; \underset{\sim}{\theta}_o) \, M(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}_o) \to - I_k \text{ en probabilité.}$$

Enfin, encore d'après les hypothèses 5.4. et 5.6., il résulte que

$$\left| I^{-1}(n ; \underset{\sim}{\theta}_o) \, V^{(1)}(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}, \underset{\sim}{\theta}_o) \right| < \delta^2 \, M^{(1)}$$

où $M^{(1)}$ est un vecteur de constantes et $\| \underset{\sim}{\theta} - \underset{\sim}{\theta}_o \| < \delta$ où $\delta > 0$ est arbitraire.

Donc, on a

(5.2.9.) $(\underset{\sim}{\theta} - \underset{\sim}{\theta}_o)^T \, I^{-1}(n ; \underset{\sim}{\theta}_o) \, \underset{\sim}{l}(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}_o) \leqslant$

$$\leqslant - \| \theta - \theta_o \|^2 + \delta^2 M^{(2)} \| \underset{\sim}{\theta} - \underset{\sim}{\theta}_o \| \leqslant - \delta^2 M^{(3)} < 0 ,$$

où $M^{(2)}$ et $M^{(3)}$ sont des constantes positives.

Donc, puisque $\underset{\sim}{l}(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta})$ est dérivable en $\underset{\sim}{\theta}$ d'après le lemme 2 de [3] il s'ensuit qu'il existe une racine $\hat{\underset{\sim}{\theta}}_n(\underset{\sim}{X}^{(n)})$ telle que

$$\underset{\sim}{l}(\underset{\sim}{X}^{(n)} ; \hat{\underset{\sim}{\theta}}_n(\underset{\sim}{X}^{(n)})) = 0 \quad \text{et} \quad \| \hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_o \| < \delta .$$

$\delta > 0$ étant arbitraire, on a que $\hat{\underset{\sim}{\theta}}_n(\underset{\sim}{X}^{(n)})$ est un estimateur convergent de $\underset{\sim}{\theta}_o$.

Pour démontrer que $\hat{\underset{\sim}{\theta}}_n(\underset{\sim}{X}^{(n)})$ est asymptotiquement efficace, posons $A^2(n ; \underset{\sim}{\theta}) = I(n ; \underset{\sim}{\theta})$.

On a

(5.2.10) $$0 = \underset{\sim}{l}(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}_o) + M(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}_o)(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_o) + \frac{1}{2} V^{(1)}(\underset{\sim}{X}^{(n)} ; \hat{\underset{\sim}{\theta}}_n , \underset{\sim}{\theta}_o)$$

ou

$$0 = A^{-1}(n ; \underset{\sim}{\theta}_o) \, \underset{\sim}{l}(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}_o) + A^{-1}(n ; \underset{\sim}{\theta}_o) M(X^{(n)} ; \underset{\sim}{\theta}_o) .$$

$$. A^{-1}(n ; \underset{\sim}{\theta}_o) A(n ; \underset{\sim}{\theta}_o)(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_o) + \frac{1}{2} A^{-1}(n ; \underset{\sim}{\theta}_o) V^{(1)}(\underset{\sim}{X}^{(n)} ; \hat{\underset{\sim}{\theta}}_n , \underset{\sim}{\theta}_o) .$$

Mais $A^{-1}(n ; \underset{\sim}{\theta}_o) M(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}_o) A^{-1}(n ; \underset{\sim}{\theta}_o) \to - I_k$ en probabilité et $\frac{1}{2} A^{-1}(n ; \underset{\sim}{\theta}_o) V^{(1)}(\underset{\sim}{X}^{(n)} ; \hat{\underset{\sim}{\theta}}_n , \underset{\sim}{\theta}_o) \to 0$ en probabilité.

Donc, on a

(5.2.11.) $$\left| A^{-1}(n ; \underset{\sim}{\theta}_o) \, \underset{\sim}{l}(\underset{\sim}{X}^{(n)} ; \underset{\sim}{\theta}_o) - A(n ; \underset{\sim}{\theta}_o)(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_o) \right| \to 0$$

en probabilité, ce qui démontre la proposition.

Remarque. — Les hypothèses sous lesquelles la proposition a été démontrée peuvent être affaiblies. Par exemple, l'existence des dérivées partielles troisièmes peut se remplacer par une condition de type de Lipschitz

comme dans le chapitre 4. Une autre affaiblissement apparent sera de considérer un ensemble de fonctions $H_{ijl}(\underset{\sim}{X}^{(n)})$ chacune satisfaisant à l'hypothèse 5.4. au lieu d'une seule $H_n(\underset{\sim}{X}^{(n)})$.

5.3. Estimation pour des processus de Markov

Comme dans le chapitre 4 , nous pouvons démontrer les propriétés optimales des estimateurs du maximum de vraisemblance lorsqu'il s'agit du cas de plusieurs paramètres dans tous les cas traités. A titre d'exemple, nous allons montrer ici seulement le cas du processus de Markov à temps discret correspondant au paragraphe 4.3.1.

Les hypothèses qu'on fait ici sont les mêmes que dans le chapitre 1 (hypothèses 1.1. - 1.6.) .

PROPOSITION 5.3.1. — *Sous les hypothèses* 1.1. - 1.6., *l'équation de de vraisemblance a, avec une probabilité qui tend vers* 1 , *quand* $n \to \infty$, *une racine* $\underset{\sim}{\hat{\theta}}_n(\underset{\sim}{X}^{(n)})$ *dans un voisinage de la vraie valeur du paramètre* θ_o .

(ii) $\underset{\sim}{\hat{\theta}}_n(X^{(n)})$ *est un estimateur convergent de* $\underset{\sim}{\theta}_o$

(iii) $\underset{\sim}{\hat{\theta}}_n(\underset{\sim}{X}^{(n)})$ *est asymptotiquement efficace.*

(iv) $\underset{\sim}{\hat{\theta}}_n(\underset{\sim}{X}^{(n)})$ *suit asymptotiquement la loi normale d'espérance mathématique* $\underset{\sim}{\theta}_o$ *et de matrice de covariance* $[(n-1)\ \Gamma(\theta_o)]^{-1}$.

DÉMONSTRATION. — La démonstration de (i) et (ii) se fait comme dans le cas général ou comme le cas d'un paramètre. Pour montrer (iii) l'équation de vraisemblance s'écrit

$$(5.3.1.)\quad \frac{\partial \log L_n(\underset{\sim}{X}^{(n)};\ \underset{\sim}{\theta})}{\partial \theta_i} = \left(\frac{\partial \log L_n}{\partial \theta_i}\right)_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_o} + \sum_{j=1}^{k}(\theta_j - \theta_j^\circ)\left(\frac{\partial^2 \log L_n}{\partial \theta_i\, \partial \theta_j}\right)_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_o} - \frac{1}{2}\sum_{j,l=1}^{k}(\theta_k - \theta_j^\circ)(\theta_i - \theta_i^\circ)\left(\frac{\partial^3 \log L_n}{\partial \theta_i\, \partial \theta_j\, \partial \theta_l}\right)_{\underset{\sim}{\theta}}$$

pour $i = 1,\ldots,k$.

Ou en notation matricielle

(5.3.2.) $\underset{\sim}{l}(\underset{\sim}{X}^{(n)}; \underset{\sim}{\theta}) = \underset{\sim}{l}(\underset{\sim}{X}^{(n)}; \underset{\sim}{\theta}_o) + M(\underset{\sim}{X}^{(n)}; \underset{\sim}{\theta}_o)(\underset{\sim}{\theta} - \underset{\sim}{\theta}_o) + V^{(1)}(\underset{\sim}{X}^{(n)}; \underset{\sim}{\theta}, \underset{\sim}{\theta}_o)$.

Ainsi, en remplaçant $\underset{\sim}{\theta}$ par $\underset{\sim}{\hat{\theta}}_n$, racine du système d'équations de vraisemblance

(5.3.3.) $$\frac{1}{n-1}\left[\underset{\sim}{l}(\underset{\sim}{X}^{(n)}; \underset{\sim}{\theta}_o) + M(\underset{\sim}{X}^{(n)}; \underset{\sim}{\theta}_o)(\underset{\sim}{\hat{\theta}}_n - \underset{\sim}{\theta}_o) + V^{(1)}(\underset{\sim}{X}^{(n)}; \underset{\sim}{\hat{\theta}}_n, \underset{\sim}{\theta}_o)\right] = 0$$

ou

$$(n-1)^{-1}\,\Gamma^{-1}(\underset{\sim}{\theta}_o)\,\underset{\sim}{l}(\underset{\sim}{X}^{(n)}; \underset{\sim}{\theta}_o) + \Gamma^{-1}(\underset{\sim}{\theta}_o)\,\frac{1}{(n-1)}\,M(\underset{\sim}{X}^{(n)}; \underset{\sim}{\theta}_o)(\underset{\sim}{\hat{\theta}}_n - \underset{\sim}{\theta}_o) +$$

$$+\,\Gamma^{-1}(\underset{\sim}{\theta}_o)\,.\,\frac{1}{n-1}\,V^{(1)}(\underset{\sim}{X}^{(n)}; \underset{\sim}{\hat{\theta}}_n, \underset{\sim}{\theta}_o) = 0\ .$$

ou

(5.3.4.) $$\sqrt{(n-1)^{-1}\,\Gamma^{-1}(\underset{\sim}{\theta}_o)}\,\underset{\sim}{l}(\underset{\sim}{X}^{(n)}; \underset{\sim}{\theta}_o) - \sqrt{(n-1)\,\Gamma(\underset{\sim}{\theta}_o)}\,(\underset{\sim}{\hat{\theta}}_n - \underset{\sim}{\theta}_o) + \underset{\sim}{\xi}_n = 0$$

où $\underset{\sim}{\xi}_n \to 0$ en probabilité, d'où le résultat.

Enfin, puisque $\sqrt{(n-1)^{-1}\,\Gamma^{-1}(\underset{\sim}{\theta}_o)}\,\underset{\sim}{l}(\underset{\sim}{X}^{(n)}; \underset{\sim}{\theta}_o)$ suit asymptotiquement la loi normale réduite, il résulte également que $\sqrt{(n-1)\,\Gamma(\underset{\sim}{\theta}_o)}\,(\underset{\sim}{\hat{\theta}}_n - \underset{\sim}{\theta}_o)$, suit asymptotiquement la même loi gaussienne, ce qui démontre (iv) .

Remarque 1. — Comme dans le cas d'un paramètre, l'existence des dérivées partielles troisièmes peut se remplacer par une condition de type de Lipschitz .

Remarque 2. — Dans le cas des processus de Markov à temps continu, le problème peut se ramener au problème correspondant au processus inclus.

Par conséquent, le problème des tests d'hypothèses concernant plusieurs paramètres se traite comme dans le cas des processus de Markov à temps discret, en considérant le processus inclus et les résultats des chapitres 1 et 2 peuvent s'appliquer dans ce cas.

*

* *

BIBLIOGRAPHIE

[1] ADHIKARI (B.P.) .– Thèse Doctorat ès Sciences Mathématiques, Université de Paris, (1958).

[2] ADHIKARI (B.P.) & JOSHI (D.D.). – Distance, discrimination et résumé exhaustif, *Publ. Inst. Statist. Univ. Paris*, 5 (1956), 57-74.

[3] AITCHISON (J.) & SILVEY (S.D.). – Maximum likelihood estimamation of parameters subject to restraints, *Ann. Math. Statist., 29* (1958), 813-828.

[4] AITKEN (A.C.). – *Determinants and matrices,* Oliver and Boyd, London (1952).

[5] ANDERSON (T.W.). – *An introduction to multivariate statistical analysis,* John Wiley, New York (1958).

[6] ANDERSON (T.W.) & GOODMAN (L.A.). – Statistical inference about Markov chains, *Ann. Math. Statist., 28* (1957), 89-110.

[7] ARTIN (E.). – *The gamma function,* Holt, Reinhart and Wiston, New York (1962).

[8] BAHADUR (R.R.). – On asymptotic efficiency of tests and estimates, *Sankhyā , 22* (1960), 229-252.

[9] BAHADUR (R.R.). – On Fisher's bound for asymptotic variances, *Ann. Math. Statist., 35* (1964), 1545-52.

[10] BARTLETT (M.S.). – Frequency goodness of fit test for probability chains, *Proc. Cambridge Philos. Soc., 47* (1951), 86-95.

[11] BARTLETT (M.S.). – Approximate confidence intervals II - More than one unknown parameter, *Biometrika, 40* (1953), 306-317.

[12] BASU (D.). – Choosing between two simple hypotheses and the criterion of consistency, *Proc. Nat. Inst. Sci. India, 19* (1953), 841-849.

[13] BILLINGSLEY (P.). – *Statistical inference for Markov Processes,* University Chicago Press (1961).

[14] BILLINGSLEY (P.). — Statistical methods in Markov chains, *Ann. Math. Statist., 32* (1961), 12-40.

[15] BILLINGSLEY (P.). — Lévy-Lindeberg theorem for martingales, *Proc. Amer. Math. Soc., 12* (1961), 788-792.

[16] BRIEMAN (L.). — An elementary theorem concerning stationary ergodic process, *Ann. Math. Statist., 29* (1958), 592-599.

[17] BUI TRONG LIEU. — Estimation pour des processus de Markov, *Publ. Inst. Statist. Univ. Paris, 11* (1962), 72-189.

[18] CHAPMAN (D.) & ROBBINS (H.). — Minimum variance estimation without regularity conditions, *Ann. Math. Statist., 22* (1951), 581-586.

[19] CRAMER (H.). — *Mathematical methods of statistics,* Princeton University Press, Princeton (1946).

[20] DOOB (J.L.). — *Stochastic processes,* John Wiley, New York (1953).

[21] DUGUE (D.). — Application des propriétés de la limite au sens du calcul des probabilités à l'étude des diverses questions d'estimation, *Journal de l'Ecole Polytechnique, 3* (1937), 305-372.

[22] FRECHET (M.). — *Généralités sur les probabilités éléments aléatoires,* Gauthier-Villars, Paris (1951).

[23] GRENANDER (U.). — Stochastic processes and statistical inference, *Ark. Mat., 1* (1950), 195-277.

[24] HALMOS (P.R.). — *Measure theory,* Van Nostrand, New York (1950).

[25] HALMOS (P.R.) & SAVAGE (L.A.). — Application of Radon-Nikodym theorem to the theory of sufficient statistics, *Ann. Math. Statist., 20* (1949), 225-240.

[26] IBRAGIMOV (I.A.). — A central limit theorem for a class of dependant random variables, *Theor. Probability Appl., 8* (1963), 83-89.

[27] ISSACSON (S.L.). — On the theory of unbiased tests of simple statistical hypotheses specifying the value of two or more parameters, *Ann. Math. Statist., 22* (1951), 217-234.

[28] KAKUTANI (S.). — On equivalence of infinite product measures, *Ann. of Math., 49* (1948), 214-224.

[29] KALLIANPUR (G.). — Von Mises functionals and maximum likelihood estimation, *India Statistical Society, Calcutta* (1963), 137-146.

[30] KALLIANPUR (G.) & RAO (C.R.). — On Fisher's lower bound to asymptotic variance of a consistent estimate, *Sankhyā., 15* (1955), 331-342.

[31] KRAFT (C.). — Some conditions for consistency and uniform consistency of statistical procedures, *Univ. California Publ. Statist., 2, n° 6* (1955), 125-142.

[32] LECAM (L.). — On some asymptotic properties of maximum likelihood estimates and related Bayes estimates, *Univ. California Publ. Statist., 1, n° 11* (1953), 277-330.

[33] LEHMANN (E.L.). — *Testing statistical hypotheses,* John Wiley, New York (1959).

[34] LEHMANN (E.L.). — On the non verifiability of certain parametric functions, *Theor. Probability Appl., 10* (1965), 689-691.

[35] LINNIK (Ju.V.). — Cours à l'Institut de Statistique de l'Université de Paris (1967).

[36] LOEVE (M). — *Probability theory,* 3ème éd., Van Nostrand, New York (1963).

[37] MIHRAM (C.A.) & STACY (E.W.). — Parameter estimation for a generalized gamma distribution, *Technometrics,* 7 (1965), 349-358.

[38] NEVEU (J.). — *Bases mathématiques du calcul des probabilités,* Masson, Paris (1964).

[39] NEYMAN (J.). — *Optimal asymptotic tests of composite statistical hypotheses* - Probability and statistics - The Harald Cramer Volume, U. Grenander, John Wiley, New York (1959), 213-234.

[40] NEYMAN (J.) & PEARSON (E.S.). — Contributions to the theory of testing statistical hypotheses, parts II, III, *Statistical Research Memoirs, 2* (1938), 25-57.

[41] NEYMAN (J.) & SCOTT (E.L.). — On the use of C (α) optimal tests of composite hypotheses, *Bull. Inst. Internat. Statist., 41* (1965), 477-497.

[42] PARZEN (E.). — On uniform convergence of families of sequences of random variables, *Univ. California Publ. Statist., 2, n° 2* (1954), 23-54.

[43] POTI (S.J.) & RAO (C.R.). — On locally most powerful tests when the alternatives are one-sided, *Sankhyā*, 7 (1946), 439.

[44] RAO (C.R.). — Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, *Proc. Cambridge Philos. Soc., 44* (1948), 50-57

[45] RAO (C.R.). — The utilization of multiple measurements in problems of biological classification, *J. Roy. Statist. Soc. Ser. B, 10* (1948), 159-203.

[46] RAO (C.R.). — Asymptotic efficiency and limiting information, *Proc. 4ème Berkeley Symposium Math. Statist. and Prob., 1* (1961), 531-545.

[47] RAO (C.R.). — Apparent anomalies and irregularities in maximum likelihood estimation, *Sankhyā, 24* (1962), 73-102.

[48] RAO (C.R.). — Efficient estimates and optimum inference procedures in large samples, *J. Roy. Statist. Soc. Ser. B, 24* (1962), 46-72.

[49] RAO (C.R.). — Criteria of estimation in large samples, *Sankhyā, 25* (1963), 189-206.

[50] RAO (C.R.). — *Linear Statistical inference and its applications*, John Wiley, New York (1965).

[51] STACY (E.W.). — A generalization of gamma distribution, *Ann. Math. Statist., 33* (1962), 1187-1192.

[52] WALD (A.). — Some examples of asymptotically most powerful tests, *Ann. Math. Statist., 12* (1941), 396-408.

[53] WALD (A.). — Asymptotically shortest confidence intervals, *Ann. Math. Statist., 13* (1942), 127-137.

[54] WALD (A.). — Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Trans. Amer. Math. Soc.*, *54* (1943), 426-482.

[55] WALD (A.). — Asymptotic properties of maximum likelihood estimate of an unknown parameter of a discrete stochastic process, *Ann. Math. Statist.*, *19* (1948), 40-46.

[56] WHITTLE (P.). — Some distributions and moment formulae for Markov chains, *J. Roy. Statist. Soc. Ser. B*, *17* (1955), 235-242.

[57] WILKS (S.S.) & DALY (J.F.). — An optimum property of confidence regions associated with the likelihood function, *Ann. Math. Statist.*, *10* (1939), 225-235.

[58] WOLFOWITZ (J.). — The method of maximum likelihood and Wald's theory of decision functions, *Indag. Math.*, *115* (1953), 114-119.

[59] WOLFOWITZ (J.). — Asymptotic efficiency of maximum likelihood estimator, *Theor. Probability Appl.*, *10* (1965), 247-260.

Reçu le 29 janvier 1968

Institut de Statistique
de l'Université de Paris.

IN n° 8 - 666050 - 5

IMP. S.P.I.T. - Villeneuve St Georges